A LINEAR SUBSPACE APPROACH TO BURST

COMMUNICATION SIGNAL PROCESSING

DISSERTATION
Daniel Erik Gisselquist
Major, USAF

AFIT/DS/ENG/04-02

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

A LINEAR SUBSPACE APPROACH TO BURST
COMMUNICATION SIGNAL PROCESSING

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Electrical Engineering

Daniel Erik Gisselquist, B.S.C.S., B.S.M.S., M.S.C.E.

Major, USAF

March, 2004

AFIT/DS/ENG/04-02
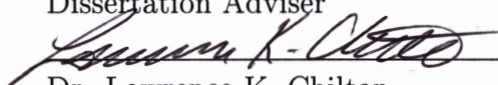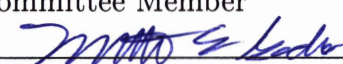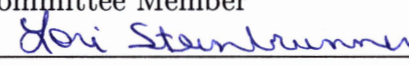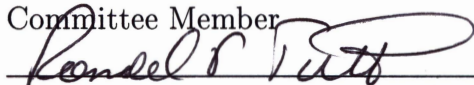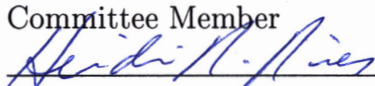
A LINEAR SUBSPACE APPROACH TO BURST COMMUNICATION SIGNAL PROCESSING

DISSERTATION

Daniel Erik Gisselquist, B.S.C.S., B.S.M.S., M.S.C.E.

Major, USAF

Approved:

Col. Donald R. Kitchen
Dissertation Adviser

8 MARCH 04
Date

Dr. Lawrence K. Chilton
Committee Member

8 Mar 04
Date

Maj. Matthew E. Goda
Committee Member

8 MAR04
Date

Dr. Lori A. Steinbrunner
Committee Member

8 Mar04
Date

Dr. Ronald F. Tuttle
Committee Member

8 MAR04
Date

Dr. Heidi R. Ries
Dean's Representative

12Mar04
Date

Accepted:

Robert A. Calico, Jr.
Dean, Graduate School of Engineering and Management

12March04
Date

AFIT/DS/ENG/04-02

Abstract

The initial intent of this research was to develop a better burst communication signal detector for high interference environments. Current burst detectors, whether of the radiometric or cyclostationary variety, are ineffective in these environments due to their lack of resistance to burst interference. This lack of resistance can be traced to the assumptions underlying the development of each of these detectors. Radiometers follow a development for stationary signals, which assumes that the statistics of a signal are independent of time. This assumption is not valid for communications signals. Cyclostationary feature detectors, on the other hand, assume that the statistics are periodic with time. While this is true of many communications signals, it is not true of burst communications signals. Therefore, this work derives, from first principles, a linear subspace approach to deriving signal processing algorithms for burst communications signals. Unlike stationary or cyclostationary approaches, this method assumes that the signal of interest is finite in length, rather than infinite. This new approach is then applied to three different application areas: Binary Phase Shift Keyed (BPSK) signal demodulation, Time Difference of Arrival (TDOA) estimation, and finally to the original signal presence detection problem. Improvements demonstrated in each area validate this method.

Given that statistical detection techniques require estimates of the unknown parameters, the first application of this new approach was estimating the unknown message symbols. When applied to BPSK signals, this approach led to two new results. The first is the derivation and specification of an optimal, minimum mean square error, linear filter appropriate for demodulating burst communications received by one or more sensors. Performance bounds were then calculated for both this new filter as well as an arbitrary demodulator. The second new result is a method of demodulator performance prediction capable of predicting the degradation resulting from a non–optimal receiver configuration.

iv

Other parameters, such as the TDOA between two sensors, are also well estimated by this approach. This is demonstrated by developing first a maximum likelihood TDOA estimator for burst communications signals, and then a Cramér–Rao bound appropriate for bounding the performance of this estimator. Unlike other developments, these methods are derived under strong interference assumptions. The result is an estimator that outperforms other "optimal" TDOA estimators in simulated interference environments by 15–20% for white noise, and by up to 70% in one wideband interference environment. While this demonstrated performance does not achieve the new Cramér–Rao bound, it compares favorably to it.

Returning to the detection problem, applying this new approach resulted in a new class of signal detector: the *cyclic ratio detector*. Unlike previous selective detectors whose false alarm rates approach 100% in strong burst interference, this new detector maintains a low false alarm rate even in the presence of burst interference. This allows it to distinguish high energy interference bursts from either high or low energy signal bursts—a capability not found in other burst detectors.

All of these algorithms are simple consequences of the new linear subspace approach to burst communications signal processing.

Acknowledgements

In the spirit of giving credit where credit is due, this is my opportunity to thank those who have had significant impacts on my efforts over the past three years. Apart from these people, this research would not be possible. Indeed, in many cases, it would never have gotten off the ground.

First, I owe my adviser, Col. Kitchen, a word of thanks. He was the first adviser in the department daring enough to work on a problem with potentially significant military application. He has taught me, more than anyone else at AFIT, what true research is. This document owes much to his tutoring, instructing, and mentoring. Thus, to my adviser, thank you.

My committee has certainly been a help to me as well. I am very thankful for their willingness to spend time working with me, sometimes at a moments notice, as I struggled to refine these ideas. In particular, Dr. Tuttle arranged crucial sponsor support just when I needed it. Dr. Chilton spent countless hours with me examining my proofs and derivations. His personal approach, that of trying to improve things rather than just finding mistakes, made him a joy to work with. Next, Maj. Goda, Dr. Ries, and Dr. Steinbrunner all made themselves available, easy to work with, and a great help reviewing and finalizing this work. So, to my committee, thank you.

As to sponsors, this work has generated a large list of official and unofficial sponsors. Indeed, I owe a debt of gratitude to all of those who have helped me. Of these, I would like to mention Mr. Hanus, and Mr. Apisa for the extensive amount of their own time that they dedicated to this project—often at the expense of whatever else they were doing. Mr. Stephens also opened up his personal library to my perusal and borrowing, and Mrs. Hopper continually found audiences for me to present my work to for review. Therefore, my thanks goes out to these and other sponsors who supported me in this endeavor.

Finally, if I have missed anyone, please accept my apologies. I am grateful for your support, as well as the support of many others who are not mentioned here.

Daniel Erik Gisselquist

Personal Acknowledgements

Others have helped me through this endeavor as well, in perhaps a different way, and I would like to mention them here. Chief among them is my God, as His help was obvious to me during this research, and well appreciated. Two examples will illustrate this. First, on those nights when I found myself unable to solve some portion of this problem, I would pray for help before going to bed. Without fail, help came in the morning. Problems that were impossible the night before, became simple in the morning. If this was not God's help, I do not know what it was.

> Some trust in chariots, and some in horses: but we will remember the name of the LORD our God. —*Psalms 20:7*

The second reason why I owe my thanks to God can be understood through my experiences with other tasks. I have seen missions both succeed and fail, and many that have failed have done so as a result of providential circumstances that could not be controlled. Perfect preparation was not the problem.

> The horse is prepared against the day of battle: but safety is of the LORD.—*Proverbs 21:31*

But this research effort was different. Every circumstance in this effort has worked to my advantage. Therefore my thanks goes first, and foremost, to my God who arranged every circumstance to my advantage.

Last, but certainly not least, I need to thank my family for their support. First, for my wife, who got excited with me about things she never fully understood and helped me regardless. Second, for my children, who constantly called me from work to enjoy some time of play. Thanks, kids, for playing when we could and for understanding when I said no. So to a family who will never fully understand their contribution, thank you.

Daniel Erik Gisselquist

Table of Contents

List of Figures

List of Tables

List of Symbols

xv

List of Abbreviations

Abbreviation		Page
BER	Bit Error Rate	70
BPSK	Binary Phase Shift Keyed	3
CDF	Cumulative Distribution Function	107
CPM	Coherent Phase Modulation	7
CPS	Cycles per Sample	114
CRB	Cramér–Rao Bound, or Cramér–Rao Lower Bound	30
FDOA	Frequency Difference of Arrival	184
FFT	Fast Fourier Transform	87
FIM	Fisher Information Matrix	73
GCC	Generalized Cross Correlation	25
ISI	Intersymbol Interference	21
LPF	Lowpass Filter	64
MCRB	Modified Cramér–Rao Bound	72
MF	Matched Filter	115
ML	Maximum Likelihood	218
MLE	Maximum Likelihood Estimator	26
MMSE	Minimum Mean Square Error	4
MSK	Minimum Shift Keying	7
OQPSK	Offset Quadrature Phase Shift Keying	7
PAM	Pulse Amplitude Modulation	7
PSD	Power Spectral Density	9
QAM	Quadrature Amplitude Modulation	7
QPSK	Quadrature Phase Shift Keying	7
SCF	Spectral Correlation Function	15
SNR	Signal to Noise Ratio	29
SPECCOA	Spectral Coherence Alignment method	28
TDL	Tapped Delay Line	4
TDOA	Time Difference of Arrival	3
UMP	Uniformly Most Powerful	225

A Linear Subspace Approach to Burst Communication Signal Processing

I. Problem Introduction

When this research began, there were only two types of burst signal detectors in the literature. The first type of detector, a radiometer or energy detector, had been well studied for over thirty years [67]. Since it used the energy within the signal to detect the signal's presence, it was highly susceptible to burst interference [60]. The second type of detector, a cyclostationary feature detector, was introduced to overcome these limitations [17]. Yet while cyclostationary detectors offer more selectivity than their energy detector counterparts, long observation times are necessary for selectivity. These deficiencies became the reason for re–examining burst signal processing.

As a result, this study began by focusing on the assumptions underlying each detection method. The most common assumption, that the statistics of the waveform under examination do not change with time, leads to *stationary* signal processing. Since stationary signal processing underlies all measurements of signal frequency usage, it is central to both system design, performance analysis, and frequency allocation [4]. This stationary assumption has also been central to statistical algorithm development. Without it, a waveform cannot be analyzed in a time independent fashion, and the time when the observation is made becomes important—complicating algorithm development. Indeed, the analytical problems solved by this one assumption have simplified algorithm development for years.

The biggest problem with stationary signal processing comes from common communications signals which have underlying periodicities not accounted for if the

1

observation time is irrelevant. Indeed, the statistics of these waveforms often vary in a periodic fashion. While this periodic variation has been known since Nyquist [46], only advances in signal processing over the last three decades have begun to take advantage of these underlying periodicities. In particular, new techniques that leverage these *cyclostationary* properties, as they are called, have outperformed older techniques based upon the stationary assumption alone. The reason for the success of these new techniques is quite simple: cyclostationary statistics model more of the properties inherent in digital communication signals than stationary statistics alone [21].

Yet cyclostationary techniques, while superior, are often difficult to derive. The primary difficulty is that, by definition, a cyclostationary signal must have an infinite length. Therefore, the application of cyclostationary signal processing techniques has been typically confined to long duration signals.[1] A second difficulty comes from the fact that cyclostationary techniques theoretically achieve complete immunity to noise and interference when applied to infinite length signals. Thus they are commonly derived for benign interference environments, and then applied in high interference environments.

Despite these difficulties, previous authors have exploited cyclostationary properties in digital communications signals to demonstrate improved demodulators [2], time difference of arrival estimators [61, 23], and signal presence detectors [17]. In addition, certain approximations to these methods have resulted in *signal selective* methods. The desire for optimality, together with a desire for improved interference resistance, drives the search for better methods.

All these problems developing cyclostationary algorithms are resolved, however, when the signal is described using classical linear subspace theory instead of either stationary or cyclostationary signal models. This dissertation will demon-

---

[1]See Sec. 2.1 for a more detailed description of the problems associated with applying classical cyclostationary approaches to burst communication signal processing.

strate this by showing how linear subspace theory can be used to derive optimal signal processing algorithms. By using a linear subspace approach, known cyclostationary features will be shown to be consequences of the method, rather than the definition of it. Further, building interference resistance into an algorithm is a natural consequence of the linear subspace framework.

1.1 Specific Application Areas of Interest

To show the validity of a linear subspace approach to communications signal processing, it is applied to three application areas within burst signal processing: Binary Phase Shift Keyed (BPSK) signal demodulation, Time Difference of Arrival (TDOA) estimation, and signal presence detection. As this research will demonstrate, this approach confirms known optimal algorithms when it is applied under previously studied conditions. What makes this approach special are the new signal processing techniques derived under more difficult conditions, such as burst signals in high interference environments. In particular, each of the signal processing techniques developed here will exploit knowledge of the noise covariance, a parameter describing the nature of the interference environment, which may be estimated when the signal is absent and then used when the signal is present. This covariance is then used to limit the impact of the interference in each of the three application areas.

1.1.1 BPSK Signal Demodulation.

The first application area, *symbol estimation*, is perhaps the most fundamental[2]. Statistically, this is equivalent to estimating the signal component of the received waveform. Since a good signal estimate underlies every other application, developing this estimate needs to be pursued first. Such estimation includes receiver synchronization, filtering and channel equalization. In particular, special attention is paid here to the filtering and equalization problem. The filtering problem is quite easily stated: what filter should be applied prior to

---

[2]Many texts refer to this problem as a "detection" problem. [53]. The term "symbol estimation" has been used here instead of "symbol detection" to avoid confusion with "signal presence detection" later.

symbol estimation in order to get the "best" results. Past research has shown that this optimal filter always consists of a matched filter, followed by a sampler and a Tapped Delay Line (TDL) equalizer [12]. The importance of this equalizer is highlighted by the fact that, "The revolution in data communication technology can be dated to the invention of automatic and adaptive channel equalization in the late 1960's" [28].

Yet the common approach to receiver filtering is to apply a matched filter designed under white noise conditions, followed by an adaptive equalizer to clean up any residual distortion. While this technique is appropriate for dealing with signals in a benign interference environment [30], it is inappropriate for burst signals in high interference environments. The reason for this is twofold. First, no attempt is made to remove or mitigate the interference in the demodulator. Second, in burst signal environments, the signal may not last long enough for an adaptive equalizer to converge.

An alternate approach is to design a fixed filter to achieve a Minimum Mean Square Error (MMSE) between the estimated symbols at the output of the demodulator and the values sent by the transmitter [2]. This method works by first excising any narrow–band interference from the input of the demodulator, and then by applying a fixed equalizer to the output to remove the resulting intersymbol interference. While this approach has potential, it has only been applied to baseband and complex signals. As a result, previous MMSE filters do not exploit all of the spectral redundancies found in a BPSK signal (see Sec. 3.1). Additionally, while able to predict the minimum mean square error [2], this technique has not been used to predict the performance of an arbitrary, non–optimal receiver. An improvement in either of these areas could facilitate receiver design and, in some cases, improve performance.

1.1.2 Time Difference of Arrival Estimation.   The second application area of interest is that of TDOA estimation—a fundamental parameter used to locate

a radio transmitter. In civilian contexts, locating transmitters is commonly used for search and rescue, as well as unintentional interference mitigation. In military contexts, locating such a transmitter has been historically used to find both the location of foreign spies as well as the locations of opposing military forces [36].

This problem has been so extensively studied in the literature [37, 52, 6] under stationary assumptions that standard solutions are readily available to solve it. Methods have already been developed which achieve the theoretical performance limit. As with filtering, these optimal methods first filter out interference, and then enhance the signal before estimating the TDOA.

Having said that, certain recent papers have reported cyclostationary TDOA estimates that beat this same theoretical limit under the justification that the limit was invalid since the signals in question were not truly stationary but cyclostationary in nature [24, 61]. While these estimators appear to work well, they have not been benchmarked in the open literature against optimal stationary algorithms to properly demonstrate their value. Rather, they have often been derived for white noise conditions, tested in high interference environments, and never compared to optimal stationary estimators. Further, no new limit has been developed to bound the performance of a cyclostationary TDOA estimator.

What a linear subspace framework offers is the ability to derive cyclostationary TDOA estimators in a rigorous manner. Even better, the theoretical limit for cyclostationary TDOA estimation, the Cramér–Rao bound which could not be calculated before, falls out naturally from a proper model. Both of these results, new estimators and Cramér–Rao bounds for digital communications signals, will be presented in Sec. 3.3.

1.1.3 Signal Presence Detection. This brings us back to the application area that motivated this research, signal presence detection. This is defined as determining whether or not a communications signal, burst or otherwise, is present in

a given environment.[3] Since this problem is closely tied to military communications and radar signal processing, it has been thoroughly studied for both known and unknown waveforms [43, 54].

The simplest method of determining whether or not a burst signal is present is to look for an increase in the energy in the channel. Although this method is highly effective, it suffers a lack of selectivity in changing interference environments [17]. While pre–filters can improve the performance of these methods [59], strong interference bursts can still create 100% false alarm rates, rendering all such energy detection techniques difficult to use when burst interference is present.

Cyclostationary detection techniques, on the other hand, exploit known features unique to the signals of interest, such as the baud rate or the carrier frequency, to selectively detect signals of interest only [17]. The problem with these techniques is the implicit assumption that the signal has an infinite length [20]. As a result, they lose their selectivity when applied to short bursts in a rapidly changing background environment (see Sec. 3.4.2). Indeed, Sec. 4.4 will show that, even with cyclostationary methods, strong interference bursts can still cause a 100% false alarm rate.

Dealing with rapidly changing interference environments requires more selectivity than either of these methods, energy detection and cyclostationary feature detection, provide. A new alternative, presented in Sec. 3.4.2, is a cyclic ratio detector. This new detector offers exactly the type of discrimination required in high interference environments containing both stationary and burst interference.

1.2   Problem Definition

Before beginning, some comments on problem definition and scope are in order regarding first the signal of interest, second the noise environment, and finally the

---

[3]Hereafter, the term "presence" will often be dropped and this will be referred to as simply a "detection" problem.

extensions that will be used to describe the multi–sensor reception problem required for TDOA estimation.

As far as this research is concerned, the signal of interest, $s(t)$, will be a Pulse Amplitude Modulated (PAM) signal with a finite number of symbols, N_s. Such a signal, upon reception, would produce a measurable voltage, $x(t)$, in a receiver. This voltage may be considered a random process and described by,[4]

$$x(t) = A \Re \underbrace{\left\{ \sum_{n=0}^{N_s-1} d_n \psi(t - nT_s - \tau) e^{j(2\pi f_c t + \theta)} \right\}}_{s(t)} + n(t). \tag{1}$$

The finite number of symbols assumption sets this work apart from other cyclostationary developments [18, 17, 19], making it relevant to burst communication. This PAM model is broad and general enough to describe most modern modulation types, such as BPSK signals, Quadrature Phase Shift Keyed (QPSK) signals [58], Quadrature Amplitude Modulation (QAM) signals [58], and many binary Coherent Phase Modulation (CPM) signaling types [39].[5] Every one of these modulation types will result in a received signal described by Eqn. (1) above [58].

Each of the terms in this equation may be understood in the context of how the signal is created. As an example, consider the $N_s = 12$ symbol PAM burst shown in Fig. 1. As you can see from the figure, this burst is created from a sum of pulses, $\psi(t - nT_s - \tau)$, each shown in gray. The sum of these pulses is then shown in black. These pulses are separated from each other by the symbol length, T_s, and scaled by a system gain, A, from their initial height. Further, a time delay, τ, has caused this signal to shift to the right. For simplicity in this example, the carrier frequency, f_c, and carrier phase, θ, are shown as zero.

---

[4]The function, $\Re\{\cdot\}$, is the real operator and returns the real portion of its complex argument.
[5]Appendix C extends this model to Offset QPSK (OQPSK) signals and Minimum Shift Keyed (MSK) signals.
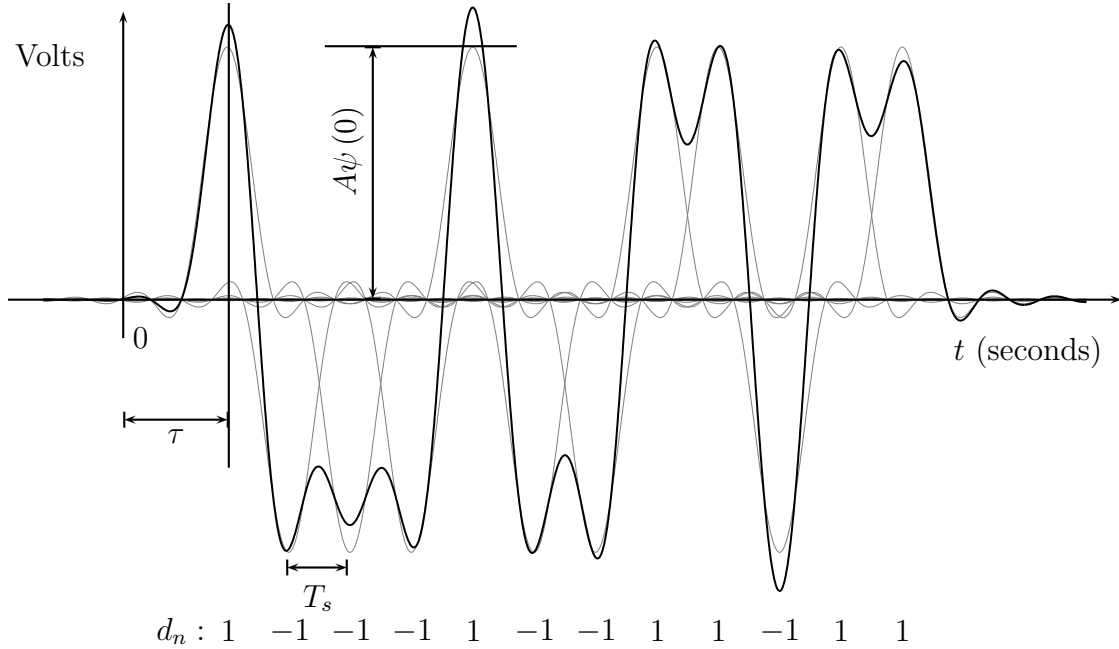
Figure 1.   A Sample PAM Signal Burst

In terms of what is known within Eqn. (1), $\psi(t)$ is the received pulse shape and assumed to be a known real function,[6] T_s is the duration of one symbol,[7] and f_c is the carrier frequency of the signal. The data symbols, d_n, are constant weights chosen according to the modulation type and the message content. As the message is unknown at the receiver, these data symbols are also assumed to be unknown under all circumstances. The rest of the signal parameters may or may not be known depending on the application area of interest. These are the phase of the carrier, θ, the time the first symbol arrives, τ, and the strength of the received signal, A. In particular, this work assumes that these parameters are known during signal demodulation and unknown otherwise.

From the standpoint of linear algebra, if we consider the set of all functions of time defined over some observation period containing the signal to be a vector

---

[6] Appendix A presents several common pulse shapes. In particular, the pulse shape shown in Fig. 1 is given by Eqn. (196) on page 187.

[7] A common related quantity, $\frac{1}{T_s}$, is the symbol rate of the communications system. It is proportional to, but not necessarily equal to, the data rate of the system.

8

space, then this burst signal lies within a subspace of that vector space. This may be seen by observation by noting that the signal is simply a linear combination of pulses whose weights are given by the d_n values. These pulses then form the basis of the linear subspace. This subspace idea will become the foundation, in Chapt. III, for a new approach to burst signal processing.

Returning the the equation that describes a received signal, the second part of this equation is the noise plus interference term, $n(t)$, often called just *noise* for short. This represents a random process that is assumed to be stationary, independent from the signal, and having a known Power Spectral Density (PSD), $S_n(f)$.[8]

These assumptions are then modified slightly for multi–sensor problems. For such problems, the noise process, $n(t)$, the carrier phase, θ, and the signal time delay, τ, may be different on each sensor. Further, the noise may have a separate power spectral density for each sensor. Two cases are treated here, depending upon the problem. For the multi–sensor demodulation problem, the carrier phase and time delay will be assumed known and the noise may be correlated across the sensors. These assumptions are different from the TDOA estimation problem, which is first simplified by assuming that the noise sources are uncorrelated and then made more complex by assuming that the carrier phase and time delay parameters are unknown for each sensor.

From these assumptions, the problem description is easy to state. For BPSK demodulation, what is the ideal way to recover the data symbols, d_n, from the received signal, $x(t)$? For TDOA estimation, what is the difference between time delay on one sensor, τ_i, and the time delay on another sensor, τ_j? Finally, for signal detection, is the received signal composed of noise alone, or does it contain some amount of signal, $A > 0$? These problems will become the focus of the next several chapters.

---

[8]See Eqn. (3) in Sec. 2.1 for a definition of this function.

1.3 Organization of this Dissertation

This dissertation is organized into five chapters. This Introduction forms the first chapter, outlining the motivation, scope and focus of this problem. The next chapter is the Background, providing a brief review of the relevant literature to these three application areas and to the problem of cyclostationary signal processing in general. This review will start with a short introduction to stationary and cyclostationary signals, and be followed by a short synopsis of previous techniques that have been applied to similar problems. The third chapter, the Theoretical Development, presents the linear subspace description of a digital communications signal. Then, classical statistical techniques are used to yield new algorithms for of the problems of filtering, TDOA estimation, and detection that exploit the periodic structure inherent in a digital communications signal. Once these algorithms are developed, the Analysis by Simulation chapter presents the performance of these new algorithms under simulated conditions. Finally, Chapt. V draws some conclusions regarding how well this new approach to burst signal processing works.

II. Background

By summarizing existing statistical models of communication signals, as well as how these models have been applied in the past, this chapter presents the need for a new look at burst signal processing. This need is shown not just for signal analysis, but for each of the three application areas of this research as well. Since this summary is quite brief, an interested reader may find additional information on each of these topics in such articles as [46, 2, 54, 19, 23, 24]. Each section within this summary presents both stationary and cyclostationary solutions. A summary at the end of each section discusses how well, or poorly, these solutions apply to the problem of burst signal processing in high interference environments.

2.1 Signal Modeling

Prior to discussing algorithm development, a discussion of the underlying approach used both to describe the signal and to derive signal processing algorithms will provide the background for understanding previous algorithms. Therefore this section describes previous models and statistics that have been used to describe both the signals of interest and the interference environment. Since both will be described in the frequency domain, the first part of this section describes the reasons for doing so, followed by the statistics of interest in the frequency domain. After that, two general approaches will be presented for deriving signal processing algorithms. The first approach treats the signal as a Gaussian stationary random process, and the second as a non–Gaussian cyclostationary process. After discussing each of these models in turn, this section will conclude with a short critique of each approach as it might be applied to burst signal processing.

For two primary reasons, all of the models discussed in this research will be discussed in the frequency domain. The first reason is that stationary signals that are correlated in time become uncorrelated in frequency for sufficiently long observation

intervals. This allows the covariance of a stationary signal to be approximated as a diagonal matrix [49]. The second reason is that unknown delay terms in the signal of interest, such as τ and θ, become complex constant multipliers in frequency, making them easier to deal with.

Within the frequency domain, the first two moments of the signal of interest, noise, and interference have been studied extensively by others. The first moment, the mean, is typically assumed to be zero for both the received signal and the interference, $\mathcal{E}\{X(f)\} = 0$.[1] The second moment is more interesting, and somewhat depends upon how $x(t)$ is defined. If $x(t)$ represents a time–limited signal, then the infinite time Fourier transform,

$$X(f) \triangleq \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt, \tag{2}$$

converges. From here, the variance in $X(f)$ can be expressed directly as $\mathcal{E}\{|X(f)|^2\}$. For signals of infinite length, this transform does not converge. An alternate method must be used to describe the variance of an infinite length signal in the frequency domain. Commonly, this is done with the power spectral density (PSD), $S_x(f)$, defined as,[2]

$$S_x(f) \triangleq \lim_{T \to \infty} \frac{1}{T} \mathcal{E}\{|X_T(f)|^2\}, \tag{3}$$

$$\text{where } X_T(f) \triangleq \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-j2\pi ft} dt. \tag{4}$$

This function describes the distribution, in power, of the signal as a function of frequency.

---

[1] $\mathcal{E}\{\cdot\}$ is used throughout this work to refer to the expected value of its argument.
[2] A related quantity, the autocorrelation function, describes the second moment of a zero mean signal in time. This function is defined for stationary signals as $R_x(\Delta t) = \mathcal{E}\{x^*(t - \frac{\Delta t}{2}) x(t + \frac{\Delta t}{2})\}$.

When signals are received across multiple sensors, correlations between these sensors may also be of interest. For arbitrary finite time signals, $x(t)$ and $y(t)$, this correlation is just $\mathcal{E}\{X^*(f)Y(f)\}$. Likewise, the cross spectral density,

$$S_{xy}(f) \;=\; \lim_{T\to\infty} \frac{1}{T}\mathcal{E}\{X_T^*(f)Y_T(f)\}, \tag{5}$$

is defined to describe this correlation for infinite length signals.[3]

One particular feature of uncorrelated signals is that their second moments sum together. Therefore, the PSD of the received signal, including both signal of interest and noise, is equal to the PSD of the signal plus the PSD of the noise [57],

$$S_x(f) \;=\; S_s(f) + S_n(f). \tag{6}$$

2.1.1 The Stationary Approach. Developing algorithms under this approach starts with the assumption that the signal is stationary, that is its probability distribution function is independent of absolute time [57]. By this definition, the only signals that are truly stationary are infinite in length. Yet hypotheses made regarding these signals are typically made surrounding some time limited observation by assuming that it has a multivariate Gaussian probability distribution. Using this approach, all received signals are treated as Gaussian random vectors, whose elements are independent in frequency [37]. Since it is Gaussian, the first two moments completely specify the distribution. The first moment, the mean, has already been assumed to be zero. The second moment in frequency, the variance, is described by the PSD of the signal. Thus all of the parameters of a stationary signal, under this model, are contained in the power spectral density.

---

[3]In general, the notation $S_{xy}(f)$ is used to represent this spectral density function between signals x and y. A similar notation, $R_{xy}(\tau)$, is used to represent their cross correlation function. This notation will be used throughout.

In particular, the PSD of a Pulse Amplitude Modulated (PAM) signal, such as the one in Eqn. (1) of interest to this research, with uncorrelated d_n values is well known [21],

$$S_s(f) = \frac{A^2}{4T_s} \left| \Psi(f - f_c) \right|^2 \mathcal{E}\left\{ |d_n|^2 \right\},\tag{7}$$

where $\Psi(f)$ is the Fourier transform of the pulse function, $\psi(t)$. Notice that the only parameters affecting this PSD are A, $\Psi(f)$, f_c, and T_s. The symbol epoch, τ, and carrier phase, θ, are not important since the statistics of stationary signals are independent of absolute time and these parameters require an absolute time reference. Some common derivations treat these two parameters, τ and θ, as random parameters instead of deterministic ones [57]. As one author puts it,

> The most common approach to modeling signals for interception studies is to ignore cyclostationarity by ... introducing a random phase variable θ uniformly distributed over one period of the cyclostationarity ... so that $x(t + \theta)$ becomes stationary. [17, p. 899]

This subtle change, from τ and θ being unknown to random, mathematically forces the theoretical probability distribution of the underlying signal to be truly stationary. This mathematical sleight of hand, however, does nothing to change the true properties of an observed signal.

The biggest problem with treating a PAM signal as a stationary process is that PAM signals are not truly stationary. Indeed, PAM signals possess significant properties not captured by this model [16, 21]. Exploiting these properties requires a different approach.

2.1.2 The Cyclostationary Approach. A second approach is to treat the signal of interest as cyclostationary. A cyclostationary signal is one whose probability distribution function is a periodic or a polyperiodic function of time [57]. As with stationary signals, any signal that meets this definition must also have an infinite

length.[4] In addition, the time of the observation is now important, since it will determine which phase of the period the observation lies within. This gives rise to unknown phase parameters, often only nuisance parameters, which need to be estimated as part of any cyclostationary signal processing algorithm.

To apply classical statistical techniques to any signal, some probability distribution function needs to be chosen to describe the signal. To date, an appropriate probability distribution has not been found for cyclostationary signals [23]. Further, the comment has been made that digital signals are not Gaussian in general [23] and therefore cannot be treated as such.[5] Without a known probability distribution, the cyclostationary algorithm designer is left to examining known moments only for properties of interest.

The advantage of cyclostationary signal processing lies in the difference between the moments of a stationary signal and those of a cyclostationary one. Cyclostationary signals have the property that particular pairs of frequencies are correlated [19, 18], while stationary signals exhibit no such correlation [49]. This correlation is called the cyclic spectral density function or the spectral correlation function (SCF) [19],

$$S_x^\alpha (f) \quad \triangleq \quad \lim_{T \to \infty} \frac{1}{T} \mathcal{E} \left\{ X_T^* \left(f - \frac{\alpha}{2} \right) X_T \left(f + \frac{\alpha}{2} \right) \right\}. \tag{8}$$

The variable α in this equation is the *cycle frequency*, or the separation in frequency between two correlated frequency pairs. It corresponds to one of the time periods found within the probability distribution function. Only man–made signals have non–zero spectral correlations when $\alpha \neq 0$. Of these man–made signals, only a finite number of values for α yield non–zero spectral correlations. Typical values for α that

---

[4]One consequence of this definition is that burst signals are neither truly stationary nor truly cyclostationary. A more appropriate description will be introduced in Sec. 3.1 that maintains the properties of interest.

[5]See Sections 3.1 and 4.1 for a demonstration of the contrary.
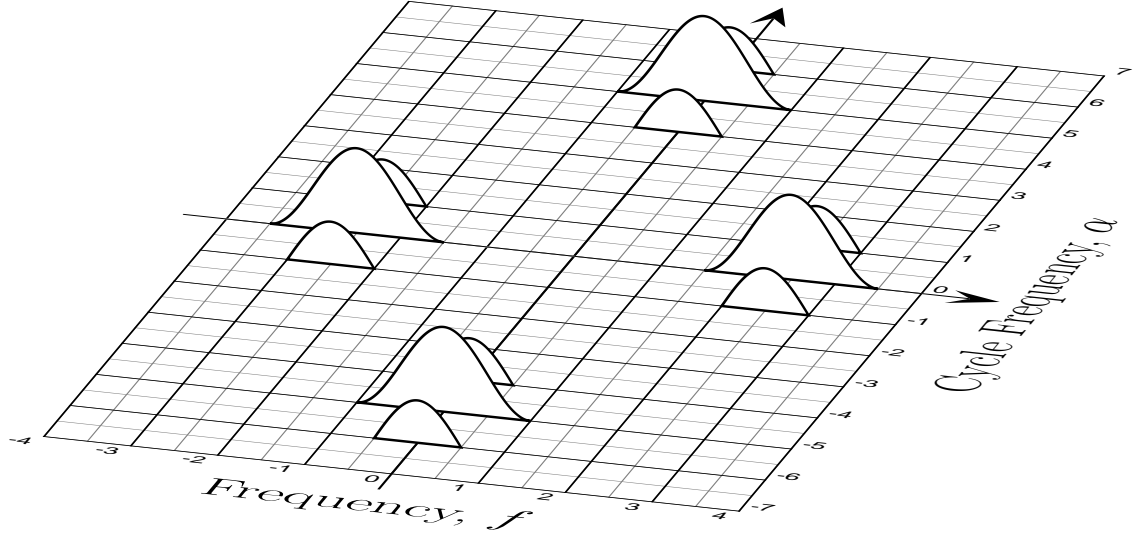
Figure 2.   The Bifrequency Plane

produce non–zero spectral correlations are zero, resulting in the PSD, multiples of the symbol rate, $\frac{k}{T_s}$, twice the carrier rate, $2f_c$, and linear combinations of these [21].

For PAM signals, the cyclic spectral density at multiples of the baud rate, $\frac{k}{T_s}$, is known to be [21]

$$S_s^{\frac{k}{T_s}}(f) \;\; = \;\; e^{-j2\pi \frac{k\tau}{T_s}} \frac{A^2}{4T_s} \Psi^* \left(f - \frac{k}{2T_s} - f_c \right) \Psi \left(f + \frac{k}{2T_s} - f_c \right) \mathcal{E} \left\{ |d_n|^2 \right\}, \quad (9)$$

for all integer values of k. In addition, PAM signals with real d_n values have a correlation surrounding twice their carrier frequency,

$$S_s^{2f_c}(f) \;\; = \;\; e^{j2\theta} \frac{A^2}{4T_s} \left| \Psi \left(f - f_c \right) \right|^2 \mathcal{E} \left\{ |d_n|^2 \right\}. \tag{10}$$

These functions are commonly plotted above a two dimensional plane with axes for the frequency, f, and the cycle frequency, α, known as the "bifrequency plane" [19]. An example bifrequency plane is shown in Fig. 2, demonstrating the cyclic correlation

16

functions for a BPSK signal having a Nyquist pulse shape, defined in Appendix A, with a data rate of one symbol per second, and a carrier frequency of 2.5 Hz.

To use these moments in practice, one simple assumption is made: only the signal of interest is correlated at a particular value of α, implying that the α chosen is non–zero. Mathematically, the cyclic spectral density of the received signal is then identically equal to the cyclic spectral density of the signal of interest, $S_x^\alpha(f) = S_s^\alpha(f)$. This assumption justifies the philosophy that estimating this value will separate the signal from all other interference in the environment. This is how signal selectivity is achieved, and it forms the basis for the common cyclostationary approach to signal processing. If the estimate of this value is non–zero, the signal is present [17]. In TDOA, the value of a similar quantity will result in the TDOA associated with the signal of interest only [23].

The two biggest disadvantages with this technique are first the noise, and second the assumption that the cyclic correlation function is unknown and needs to be estimated. The problem with the noise is that it increases the variance in any estimate of the cyclic correlation function, $\hat{S}_x^\alpha(f)$. Thus, despite the assumption that only the signal will contribute, random variations caused by noise can easily create other apparent contributions. This extra variation can be dealt with by averaging the statistic over longer and longer observations while requiring some amount of smoothness in the estimated SCF [18]. While this may be an appropriate method for working with signals of exceptionally long duration buried in noise, it is not appropriate for developing algorithms for short duration signals of interest. For short duration signals, other methods are required.

Table 1 may be helpful here to understand the difference between the stationary and cyclostationary approaches to signal processing. Under stationary assumptions, the PSD of the received signal may be measured, but the contribution due to the noise cannot be separated from the contribution due to the signal. Cyclostationary signal processing, on the other hand, focuses on the non–zero values of α where only

Table 1.   The Problem with Classical Burst Signal Analysis

Statistic		Stationary	Cyclostationary	Burst
$\alpha = 0$	$S_n(f)$	Measurable,		Measurable
	$S_s(f)$	but inseparable.		*Cannot be measured*
$\alpha \neq 0$	$S_n^\alpha(f)$	Assumed to be zero		
	$S_s^\alpha(f)$	Zero	Measurable	*Cannot be measured*

the signal contributes. By measuring these properties, and smoothing them out over time, an estimate may be made for the signal which is theoretically independent of the noise. Neither of these approaches, however, works for burst signal processing since the spectral correlation functions are only defined for infinite signals—making it impossible to measure or estimate these functions for burst signals only. The fact that the noise can be measured, however, will be discussed in more detail in the following chapter.

2.1.3   Discussion on the Fundamental Approach.   Each of these two approaches, the stationary approach and the cyclostationary signal approach, has advantages and disadvantages. While the stationary approach does not use all of the properties in the signal of interest [19], the fact that a useful probability density function can be constructed for stationary signals makes classical statistical techniques plausible. This makes the development of detection and estimation techniques for stationary signals quite straightforward. On the other hand, cyclostationary statistics account for more of the true properties within a signal. Thus, it is only reasonable to assume that techniques properly based upon this extra information would be more capable. The problem with the cyclostationary approach, however, is that it lacks a probability distribution function, which makes it difficult to apply classical statistical techniques [23]. As a result, cyclostationary algorithms have often been derived in an ad-hoc manner and validated through performance tests alone [14].

Confusing the question of which model is more appropriate is the fact that cyclostationary algorithms have rarely been compared to their optimal stationary counterparts. Two examples will illustrate this. First, single cycle detectors in [17] are only compared against standard radiometers, even though the "optimal radiometer" is well known [17, p. 903]. Second, TDOA estimates are compared in [24] and elsewhere to the Generalized Cross Correlation method rather than the optimal Eckart filter, even though the problems with the Generalized Cross Correlation method have been known for years [52]. This leaves the true relative capabilities of cyclostationary methods obscure.

For these reasons it makes sense to return to classical statistical principles, such as likelihood ratios and maximum likelihood estimation to derive optimal algorithms. The fact that these classical statistical principles have proven optimality properties associated with them makes them even more desirable. The one missing link, a workable signal model, will be presented in Sec. 3.1.

The rest of this chapter will describe standard approaches to each of three application areas—symbol estimation, TDOA estimation, and presence detection. In each case, the choice of the underlying approach will dictate the form of the solution. While a stationary approach will yield one algorithm, the cyclostationary approach will often yield a superior algorithm. This justifies the need to re–examine whether or not the cyclostationary properties contained within a burst signal may be used in a more rigorous manner to achieve even better results.

2.2   Optimal Filtering

Of the three application areas in this dissertation, optimal filtering is perhaps the oldest. The problem can be stated quite simply: what filter should be applied prior to symbol estimation in order to get the "best" results. This optimal filter is known to consist of an overwhitener, a matched filter for white noise, followed by a sampler and a Tapped Delay Line (TDL) equalizer, as shown in Fig. 3, regardless of
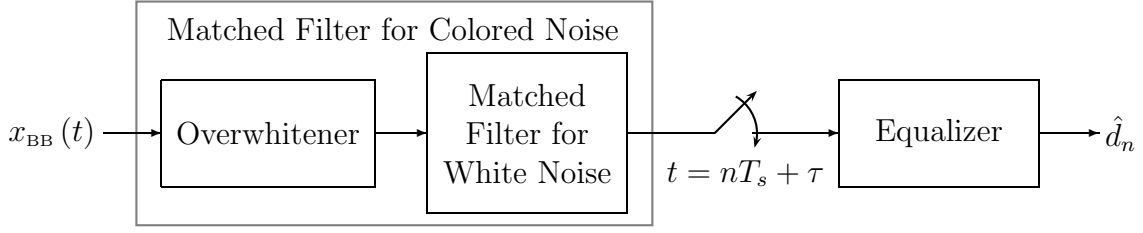
Figure 3.   The Structure of an Optimal Filter

the criteria used to determine "best" [12]. In this figure, $x_{\mathrm{BB}}(t)$ represents a signal with a zero carrier, $f_c = 0$, t denotes the time the sample is taken, and $\hat{d}_n$ is the estimated symbol at the output of the filter. The structure of this section will follow the form of this optimal filter, as first the matched filter will be discussed followed by the combination of a matched filter with an equalizer.

   2.2.1   Matched Filtering.   A matched filter is defined to be the unique filter that maximizes the signal–to–noise ratio of a known input. Common developments of this filter are presented in [58, 63, 66]. In each development, the signal is assumed to be known, such as $s(t) = A\psi(t)$, to within a constant scale factor, A. The filter, $h(t)_{\mathrm{MF}}$, that maximizes the signal–to–noise ratio in white noise is [50, 63][6]

$$h(t)_{\mathrm{MF}} \quad = \quad \psi^*(-t).\tag{11}$$

This filter can be represented in frequency as

$$H(f)_{\mathrm{MF}} \quad = \quad \Psi^*(f).\tag{12}$$

   Other developments, such as [11], extend this filter to colored noise environments. In a colored noise environment, the form of the filter that maximizes the

---

[6]White noise is a random process having a constant one–sided power spectral density, $S_n(f) = N_o$. The term N_o, here, is a constant used to specify the amount of power in the noise per unit frequency.

signal to noise ratio is

$$H\left(f\right)_{\text{MF}} \quad = \quad \frac{\Psi^{*}\left(f\right)}{S_{n}\left(f\right)}, \tag{13}$$

where $S_{n}\left(f\right)$ is the power spectral density of the noise. This filter for colored noise can be understood as a two step process. The first step whitens the noise by applying the filter

$$H\left(f\right)_{\text{Whitener}} \quad = \quad \frac{1}{\sqrt{S_{n}\left(f\right)}}, \tag{14}$$

while the second step matches the whitened output to the signal of interest. The catch is that the signal of interest no longer has the same shape—it has been distorted by the whitening filter. The signal of interest then looks like $\Psi_{w}\left(f\right) = \frac{\Psi(f)}{\sqrt{S_{n}(f)}}$. Therefore, the filter in Eqn. (13) matches to the signal as distorted by the whitening filter [66]. Since this operation can be viewed as whitening the signal twice, it is often called *overwhitening*.

While the matched filter is certainly the optimal filter when only a single pulse is transmitted, it is not necessarily the optimal filter for a communications system where multiple pulses are transmitted in succession. When multiple pulses are transmitted, the pulses may interfere with each other at the output of the filter causing Intersymbol Interference (ISI) [58]. Such interference is especially problematic in the case where $\Psi\left(f\right)$ has been designed and fixed in the transmitter, prior to the estimation of $S_{n}\left(f\right)$ and the corresponding application of the overwhitener.

2.2.2 Equalization.   Removing Intersymbol Interference (ISI) is accomplished through the use of an additional filter called an equalizer. Ericson proves, using the spectral redundancies inherent in a modulated signal, that the optimal receiving filter always consists of a matched filter followed by a TDL equalizer [12].

This equalizer can be implemented as a discrete time filter operating on the symbols that are sampled after the matched filter. Using this structure, there are two ways of applying an equalizer: fixed and adaptive.

A fixed equalizer can be designed by minimizing the mean square error between the output of the equalizer and the true symbols that were sent. Berger and Tufts present one such Minimum Mean Square Error (MMSE) filter for a pulse–amplitude modulated communication stream [2]. First, they modeled the communications signal as a sum of delayed pulses with unknown weights,

$$s(t) \;=\; \sum_{n=-\infty}^{\infty} d_n \psi(t - nT_s), \tag{15}$$

similar to Eqn. (1). Then they derive their filter from the PSD of the noise, $S_n(f)$, together with the PSD of the discrete pulse sequence $\{d_n\}_{n=-\infty}^{\infty}$, written as $S_d\left(e^{j2\pi fT_s}\right)$. This filter, they point out, is composed of both a periodic and a non–periodic component,

$$H(f)_{\text{MMSE}} \;=\; \underbrace{\left(\frac{S_d\left(e^{j2\pi fT_s}\right)}{1 + S_d\left(e^{j2\pi fT_s}\right) \dfrac{A^2}{T_s} \displaystyle\sum_{n=-\infty}^{\infty} \dfrac{\left|\Psi\left(f - \frac{n}{T_s}\right)\right|^2}{S_n\left(f - \frac{n}{T_s}\right)}} \right)}_{\substack{\text{Tapped Delay Line} \\ \text{Equalizer}}} \underbrace{\frac{\Psi^*(f)}{S_n(f)}}_{\substack{\text{Matched} \\ \text{Filter}}}. \tag{16}$$

The non–periodic component on the right hand side of Eqn. (16) is a matched filter. The term on the left side of Eqn. (16) represents the equalizer. This term is periodic with period $\frac{1}{T_s}$, making it a TDL equalizer and matching Ericson's prediction [12].

The second type of equalizer is an adaptive equalizer. These equalizers were first demonstrated by Lucky [40] in 1966 and then later studied by Haykin and many

22

others [30, 70]. They are in common use today because they require no knowledge of the interference environment in order to operate. When applied following a true matched filter, adaptive equalizers have been shown to converge to the minimum mean square error filter solution. When the initial filter is not a true matched filter, these solutions are less than optimal. Yet they can be ideal solutions to modern equalization problems where little knowledge of the noise PSD or the true signal pulse shape are available to implement the filter.

The problem with these adaptive equalizers lies in their implementation: no knowledge is used of the noise PSD or the true signal pulse shape. Without using the noise PSD to filter out narrow band interference, large amounts of interference may enter into the system. Without using the true pulse shape, signal energy is arbitrarily lost in the initial filter. While the TDL equalizer that follows may be able to reduce any resulting distortion, it cannot fundamentally compensate for poor signal to noise conditions in its input. Therefore, although practical, these systems are less than optimal.

2.2.3 Optimal BPSK Filtering. From the above discussions, it would seem that the optimal filter to use under known conditions would be a Minimum Mean Square Error filter. This filter has known optimality properties, and is defined when the pulse function for the signal and the PSD of the noise are both known. Yet Berger and Tufts' version of the MMSE filter does not exploit the extra spectral redundancies found in a BPSK signal at bandpass frequencies [2]. In particular, BPSK signals at bandpass frequencies are symmetric about their carrier frequency, and this symmetry is not exploited in a real baseband filter. As this research will show (Sec. 3.2), applying a proper filter at bandpass frequencies can improve performance under asymmetric colored noise conditions.

What may not be apparent from this discussion is that the initial filter choice lays the foundation for estimating other parameters associated with the signal, such

as synchronization parameters, TDOA estimates, and ultimately signal presence detectors. Chapt. III will highlight this connection in its development of both optimal TDOA estimators and optimal signal detectors. This underscores the criticality of this initial decision in all other burst communications processing steps.

2.3 Time Difference of Arrival Estimation

The second application area, time difference of arrival (TDOA) estimation, arises from the need to locate a transmitter by using multiple receivers. The signal of interest, arriving at each sensor, is assumed to be corrupted by additive Gaussian noise plus interference prior to TDOA estimation. For simplicity, this corruption is assumed to be statistically independent from sensor to sensor. From these assumptions, the output of the ith sensor can be expressed as [29, 37]

$$x_i(t) = A_i s(t - \tau_i) + n_i(t), \tag{17}$$

where A_i is the signal gain on the ith sensor, and τ_i is the propagation delay from the transmitter to the ith receiver. The problem of time difference of arrival estimation is to determine the parameters τ_i relative to an arbitrary time delay, τ_0, or equivalently to estimate $\tau_{di} = \tau_i - \tau_0$.[7]

2.3.1 Cross Correlation Methods.

It is commonly known that a measurement of the cross–correlation function for a zero mean signal can yield an estimate of the time delay difference between two signals. This is clearly seen by the fact that,[8]

$$R_{x_0, x_1}(\Delta t) = R_s(\Delta t - \tau_{d1}). \tag{18}$$

---

[7]A similar problem, time delay estimation, involves estimating τ_i. The difference between these two problems is whether or not the absolute, as opposed to relative, transmission time is important.

[8]$R_{xy}(\Delta t)$ is defined as the cross correlation function of the two signals, and equal to $\mathcal{E}\left\{x\left(t - \frac{\Delta t}{2}\right)^* y\left(t + \frac{\Delta t}{2}\right)\right\}$ when $x(t)$ and $y(t)$ are stationary random processes.

When the noise and the signal of interest are both white, the maximum of the cross correlation produces a remarkably clear estimate of the time difference,

$$\hat{\tau}_{d1} = \arg\max_{\Delta t} \hat{R}_{x_0,x_1}(\Delta t). \tag{19}$$

Estimating this the cross correlation, $\hat{R}_{x_0,x_1}(\Delta t)$ can be done quickly in frequency. As a result, the estimate in Eqn. (19) is often replaced by

$$\hat{R}_{x_0,x_1}(\Delta t) = \int_{-\infty}^{\infty} X_0(f) X_1^*(f) e^{j2\pi f \Delta t} df, \tag{20}$$

giving rise to the equivalent TDOA estimator,

$$\hat{\tau}_{d1} = \arg\max_{\Delta t} \int_{-\infty}^{\infty} X_0(f) X_1^*(f) e^{j2\pi f \Delta t} df, \tag{21}$$

where $X_0(f)$ and $X_1(f)$ are the time–limited Fourier transforms of $x_0(t)$ and $x_1(t)$ respectively. This method is known as the Generalized Cross Correlation (GCC) method. Although well known, it suffers problems when the signal and the noise are not white [37,52,6]. Specifically, $\hat{R}_{x_0,x_1}(\Delta t)$ may consist of multiple peaks or blurred peaks making it difficult to correctly identify the true maximum.

To circumvent this problem, a filter may be applied to the cross correlation prior to the inverse Fourier transform. This leads to a TDOA estimate of the form

$$\hat{\tau}_{d1} = \arg\max_{\Delta t} \int_{-\infty}^{\infty} H(f) X_0(f) X_1^*(f) e^{j2\pi f \Delta t} df, \tag{22}$$

where $H(f)$ is designed to create a clear cross–correlation spike.

Two choices for $H(f)$ are well known. The first is the Eckart filter and the second is the Hannan and Thompson filter, which produces the stationary Maximum

Likelihood Estimator (MLE) [37]. Both of these methods depend upon the PSD of the signal and the noise.

Of these two choices, the Eckart filter is the easiest to implement. It can be constructed entirely from values already assumed known to this research, by dividing the PSD of the signal (any sensor) by the PSDs of the noise on each of the two sensors of interest,

$$H\left(f\right)_{\text{ECKART}} \;\; = \;\; \frac{S_s\left(f\right)}{S_{n_0}\left(f\right)S_{n_1}\left(f\right)}. \tag{23}$$

Even better, this filter is provably optimal under small input signal–to–noise conditions [37].

The second choice, the Hannan Thompson filter, is said to achieve the theoretical limit in TDOA estimator performance, also known as the Cramér–Rao bound (CRB) [37].[9] This filter is a function of both the expected signal PSD, and the noise PSDs. Using Eqn. (7) for the original PSD, this estimate is,

$$\hat{\tau}_{d1,\text{ML}} \;\; = \;\; \arg\max_{\Delta t} \int_0^\infty \frac{\frac{A_0 A_1}{4T_s}\left|\Psi\left(f - f_c\right)\right|^2 \left[\frac{X_0\left(f\right)}{S_{n_0}\left(f\right)}\right]\left[\frac{X_1\left(f\right)}{S_{n_1}\left(f\right)}\right]^* e^{j2\pi f \Delta t}}{1 + \frac{A_0^2}{4T_s}\frac{\left|\Psi\left(f - f_c\right)\right|^2}{S_{n_0}\left(f\right)} + \frac{A_1^2}{4T_s}\frac{\left|\Psi\left(f - f_c\right)\right|^2}{S_{n_1}\left(f\right)}} df. \tag{24}$$

Unlike the Eckart filter, this filter requires knowledge of the unknown signal scale on both sensors, A_0 and A_1, making it difficult to implement.

---

[9]The proof that the stationary MLE achieves the CRB is made following several assumptions. When stated in 1976, experimental validation was not feasible. Experiments given in Sec. 4.3, however, demonstrate that although the stationary MLE comes close to this bound, it does not achieve it.

Assuming that the correct correlation peak is identified, the Cramér–Rao bound defining the theoretical optimal performance is, [37]

$$CRB\left(\hat{\tau}_{d1}\right) \;=\; \frac{T}{\displaystyle\int_{-\infty}^{\infty} \frac{(2\pi f)^2 \frac{A_0^2 A_1^2}{16 T_s^2} \left|\Psi\left(f - f_c\right)\right|^4}{S_{n0}\left(f\right) S_{n1}\left(f\right) \left(1 + \frac{A_0^2 |\Psi(f - f_c)|^2}{4 T_s S_{n0}(f)} + \frac{A_1^2 |\Psi(f - f_c)|^2}{4 T_s S_{n1}(f)}\right)} df}. \tag{25}$$

Knapp and Carter add the comment that this particular bound does not take into account any affects caused by correlation degradation as a result of a finite observation length.

Before leaving stationary TDOA estimation, it is important to mention the problem of subsample interpolation. In practice, TDOA estimates are created from sampled waveforms. To get the time resolution desired, subsample interpolation is required. The simplest and most common method of subsample resolution is to apply a quadratic interpolating polynomial to the output samples [5],

$$\hat{\tau}_{d1} \;=\; \hat{\hat{\tau}}_{d1} + \frac{1}{2} \frac{\hat{R}_{x0,x1}\left[\hat{\hat{\tau}}_{d1} + 1\right] - \hat{R}_{x0,x1}\left[\hat{\hat{\tau}}_{d1} - 1\right]}{2\hat{R}_{x0,x1}\left[\hat{\hat{\tau}}_{d1}\right] - \left(\hat{R}_{x0,x1}\left[\hat{\hat{\tau}}_{d1} - 1\right] + \hat{R}_{x0,x1}\left[\hat{\hat{\tau}}_{d1} + 1\right]\right)}, \tag{26}$$

where $\hat{\hat{\tau}}_{d1}$ is the initial TDOA estimate. While simple and well known, minor problems have been noticed with polynomial interpolation resulting from the fact that the true correlation peak is not a quadratic polynomial [5]. When these errors are small, polynomial interpolation is a valid means of achieving subsample resolution. When necessary, upsampling prior to quadratic interpolation may be used to minimize these errors.

2.3.2 Cyclic TDOA Estimation. There are several algorithms for measuring the TDOA of a signal using the cyclostationary properties of that signal. According to Gardner,

These new algorithms are tolerant to both interfering signals and noise, and they can outperform conventional algorithms that achieve the Cramér–Rao lower bound on variance for stationary signals because the signals considered here are nonstationary (cyclostationary) and the algorithms exploit the nonstationarity to discriminate against noise and interference. [24, p. 1197]

Two particular estimators of interest are the *Spectral Coherence Alignment method* (SPECCOA) [22, 23, 24], and Streight's cyclostationary TDOA MLE [62, 61].

SPECCOA is one of many algorithms published by Gardner and Chen on cyclic TDOA estimation [22, 23, 24]. Fong et al. comments on this technique,

One of the most successful TDOA algorithms based upon the cyclo-stationary model, which is referred to as spectral coherence alignment (SPECCOA), was derived from an ad–hoc least–squares optimization procedure. [14, p. 38]

To justify an ad–hoc algorithm, Gardner and Chen explain that,

Although classical statistical principles, such as maximum likelihood, provide an alternative approach, we have not yet found such approaches to be tractable for the non–Gaussian nonstationary models of interest. [23, p. 1177]

The algorithm itself simply involves looking at the product of two spectral correlation estimates,

$$\hat{\tau}_{d1} \;\; = \;\; \arg\max_{\tau_{d1}} \left| \int \hat{S}^{\alpha}_{x_1 x_0}\left(f\right) \hat{S}^{\alpha}_{x_0 x_0}\left(f\right) df \right|. \tag{27}$$

Although it may not be apparent, this function does depend upon τ_{d1}, since

$$S^{\alpha}_{x_1 x_0}\left(f\right) \;\; = \;\; e^{j2\pi\left(f-\frac{1}{2T_s}\right)\tau_{d1}} \frac{A_1}{A_0} S^{\frac{1}{T_s}}_{s_0}\left(f\right). \tag{28}$$

It is this dependence that SPECCOA exploits. Further, since this is one of the "most successful" cyclostationary TDOA algorithms, we consider it here [14].

One interesting thing to note about this approach is that $S_s^\alpha(f)$ is assumed to be *unknown*. Thus, accomplishing this TDOA estimate requires first estimating $S_s^\alpha(f)$ and then placing the estimate into Eqn. (27). This estimation problem, while discussed in [18], is not discussed in either of [23] or [14], leaving the full details of the experiments presented within these papers undefined. As a result, Chap. IV will present comparisons with similar single cycle estimators, rather than the actual estimators used in [23] and [14].

A second method, developed under more methodical means, is Streight's maximum likelihood TDOA estimator for low signal to noise ratio (SNR) cyclostationary signals. This estimator is given by Streight as [61, 62],

$$
\hat{\tau}_{d1} = \arg\max_{\tau_{d1}} \Re \left\{
\begin{array}{l}
\sum_\beta \int \hat{S}_{x_1 x_0^*}^\beta(f) \, S_{ss^*}^\beta(f)^* \, e^{-j\pi\beta\tau_{d1}} e^{j2\pi f \tau_{d1}} df \\[4pt]
+ \ \sum_\beta \int \hat{S}_{x_0 x_1^*}^\beta(f) \, S_{ss^*}^\beta(f)^* \, e^{-j\pi\beta\tau_{d1}} e^{-j2\pi f \tau_{d1}} df \\[4pt]
+ \ \sum_\beta \int \hat{S}_{x_1 x_1^*}^\beta(f) \, S_{ss^*}^\beta(f)^* \, e^{-j2\pi\beta\tau_{d1}} df \\[4pt]
+ \ 2\sum_\alpha \int \hat{S}_{x_0 x_1}^\alpha(f) \, S_{ss}^\alpha(f)^* \, e^{-j\pi\alpha\tau_{d1}} e^{-j2\pi f \tau_{d1}} df \\[4pt]
+ \ 2\sum_\alpha \int \hat{S}_{x_1 x_0}^\alpha(f) \, S_{ss}^\alpha(f)^* \, e^{-j\pi\alpha\tau_{d1}} e^{j2\pi f \tau_{d1}} df \\[4pt]
+ \ 2\sum_\alpha \int \hat{S}_{x_1 x_1}^\alpha(f) \, S_{ss}^\alpha(f)^* \, e^{-j2\pi\alpha\tau_{d1}} df
\end{array}
\right\}, \quad (29)
$$

where β is defined like the cycle frequency parameter, α, save that β ranges over all of those cycle frequencies where conjugation is not appropriate. As with the previous two methods, Streight also asserts that this MLE,

> ...is in fact ideal because it requires knowledge of the ideal cyclic correlation function of the SOI [Signal of Interest] a priori. In practice, this cyclic correlation function is unavailable thus requiring the ideal caveat. [61]

What is confusing from both of these references, and from similar references in [14] and [23], is the statement that these correlations cannot be known ahead of time. Why such a function cannot be guessed or approximated is never stated.

29

The issue of whether or not $S_s^\alpha(f)$ can, or cannot, be known is critical to the problem of handling bursts. If $S_s^\alpha(f)$ cannot be known, then it must be estimated and smoothed through integration over time—a commodity not available in burst signal processing. This ideal algorithm would become practical if $S_s^\alpha(f)$ could be known, if only to within a complex scale factor.

A much more difficult and related problem is that of calculating the Cramér–Rao Bound (CRB) for cyclostationary TDOA estimation. Making this problem difficult is the fact that, under the common approach to cyclostationary signal processing, no probability density function exists that can describe the cyclic properties of digital signals. Without knowing the underlying probability density functions, the gradient of the log of the probability density function, the fundamental parameter in CRB calculation, is undefined.

Two approaches have been taken to derive CRBs for cyclostationary signals.

The first approach, taken by Gardner, abandons finding the true CRB altogether, noting that

> . . . the conventional [stationary] CRLB [CRB] does not apply and,
> worse yet, the CRLB that does apply is exceedingly difficult to evaluate
> for non–Gaussian and nonstationary SOI [Signals of Interest] and SNOI
> [Signals not of Interest]. [24, p. 1193]

Instead, Gardner presents a bound based upon signal and noise energy within the bandwidth of interest—ignoring the spectral distribution of that energy [24]. This bound is,

$$CRB_{\text{GARDNER}} \;=\; \frac{3}{8\pi^2 T}\frac{1}{f_2^3 - f_1^3}\left[\frac{2N_0}{S_0} + \left(\frac{N_0}{S_0}\right)^2\right] \tag{30}$$

where

> S_0 and N_0 are the equivalent flat PSD's of the SOI and SNOI within
> the band of interest $[f_1, f_2]$, and T is the data collection time. [24, p.
> 1193]

Although ad–hoc, the resulting bound is often reasonable in white noise situations. Indeed, it appears to be the well–known CRB, given in Eqn. (25), simplified for the case where both signal and noise have white PSDs.

The problem with this bound is that it does not take into account the PSD of either the noise or of the signal. For example, a narrow band interferer near the zero frequency contributes little to the CRB in Eqn. (25) since it is scaled by $(2\pi f)^2\big|_{f=0}$, while the contribution of the same interferer increases as its frequency increases. Gardner's bound removes this dependence. Although this may be appropriate in rare circumstances, both Gardner and Streight have used this bound to benchmark the performance of algorithms in colored noise environments [24, 61] when it has never been shown to be appropriate in these environments.

Schell takes a second, more mathematical, approach to calculating CRBs using *Whittle's Theorem* [56]. Whittle's theorem states that, for stationary signals, the Fisher Information Matrix (FIM) can be expressed in terms of a diagonal matrix. Schell extends this result to cyclostationary signals by first mapping them to an underlying stationary process, and then by applying Whittle's theorem to calculate the CRBs.

This CRB is then compared with the stationary bound for TDOA estimation for AM and QAM modulation types. In particular, several features are noticed. The first is that the cyclostationary CRB converges to the stationary bound in high SNR situations. As the SNR drops, the two bounds separate by some amount, but never by more than sixty percent. Finally, Schell presents his conclusion that when a cyclostationary interferer is present, the Cramér–Rao bound is exceptionally low [56].

The problem with this approach lies in the nuisance parameters. By assuming $S_s^\alpha(f)$ is completely known Schell ignores the related but necessary problem of estimating τ and θ. Indeed, the whole concept of mapping a cyclostationary sequence into a related stationary sequence cannot be accomplished apart from knowing (or estimating) τ and θ. Finally, Schell does not demonstrate the performance of any

31

cyclic algorithms with respect to these bounds, and so his conclusions have yet to be validated [56].

Neither of these two approaches to developing a CRB take into account the spectral distribution, or coloring, of the interference. In a personal communication, Schell describes just how important this "colored noise" problem really is:

> Finally, the colored noise problem may be the most interesting problem of all in this context, because it corresponds most closely to TDOA estimation problems of practical interest. That is, in practice there are potentially numerous interfering signals that are either stationary or have totally different cycle frequencies from the signals of interest. By lumping all of this stuff-not-of-interest along with the thermal/background noise into the container called "colored noise," you can define a practical TDOA estimation problem for which the CRB might be computed (or for which a ML estimator might be derived).
> ...The mathematical difficulty of solving this resulting problem is hard to estimate. Basically, you would add all of the spatio–temporal correlations of the colored noise to the list of nuisance parameters in the problem, and then proceed to solve. The "adding to the list" part is straightforward. The "solving" part may require *super-human persistence and very deep familiarity with analytical methods of matrix computations.* [55] [Emphasis added.]

Thus, according to Schell, the colored noise problem is the general problem that is of the most practical interest. The CRB in white noise, discussed by both Schell and Gardner, should follow as a consequence of the colored noise solution. It is this colored noise problem that is of interest in this research.

Again, this highlights the need for a better approach to cyclostationary signal properties. Previous approaches, based upon second order statistics alone, have not yielded a practical CRB for TDOA estimation. With a useful model, it should be possible not only to derive a maximum likelihood estimate of the TDOA parameter, but also a CRB limiting how good any estimate might be.

$$\longrightarrow \boxed{H(f)} \longrightarrow \boxed{\left| \cdot \right|^2} \longrightarrow \boxed{\int_{t-N_s T_s}^{t} \cdot \; dt} \longrightarrow y_{\text{rad}}$$
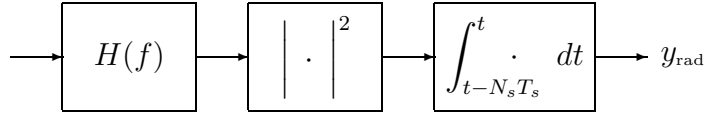
Figure 4.   A Radiometer

2.4   Presence Detection

This brings us to the final application area of interest—signal presence detection, or determining whether or not a signal is present in a noisy channel. This section presents several current methods of signal detection that have applicability to burst signal presence detection. No attempt is made to be exhaustive, and indeed several techniques have been left out. The interested reader is referred to more complete and thorough approaches, such as Scharf's text on detection and estimation [53], Van Trees classical text [68], or Casella and Berger's development of *Mathematical Statistics* [7].

The detection methods presented here can be broken down into energy detection, corresponding to stationary detection techniques, and cyclostationary detection techniques. These latter are sometimes referred to as *feature* detectors, since they detect a signal based upon unique features inherent in that signal.

2.4.1   Energy Detection Techniques.   The simplest method for detecting an unknown waveform corrupted by Gaussian noise is an energy detector, commonly known as a radiometer. This detector passes the signal first through a bandpass filter, then through a square law device, and finally integrates the result over some amount of time as shown in Figure 4. The output of this process is compared to a threshold to determine whether or not a signal was present.

33

Energy detectors for unknown signal types test the following hypothesis for some frequency band of interest [67]:[10]

$$H_0 \quad : \quad X(f) \sim \mathcal{N}(0, N_o \mathbf{I}) \tag{31}$$

$$H_1 \quad : \quad X(f) \sim \mathcal{N}(0, N_o \mathbf{I} + \sigma_s^2 \mathbf{I}), \text{ and } \sigma_s^2 > 0 \tag{32}$$

The filter $H(f)$ in this case is a bandpass filter that removes any noise from the received signal outside of the frequency band where the signal of interest should lie, potentially shapes the signal within that band, and produces a frequency vector containing the $X(f)$ tested above (Eqns. (31) and (32)). Urkowitz [67] develops this method from the hypothesis test above, noting that a vector of length $2TW$ samples is sufficient to accurately reconstruct a burst signal of duration T and bandwidth W. In his development, $H(f)$ is an ideal bandpass filter having the exact bandwidth as the signal of interest and centered at its location in frequency. The detection statistic resulting from this method is

$$y_{\mathrm{rad}}(\boldsymbol{x}) \quad = \quad \sum_{i=0}^{N-1} |X_T(f_i)|^2, \tag{33}$$

where f_i ranges over all of the frequencies that the signal occupies (see Sec. 3.1.1).

Calculating the false alarm rate for this type of detector usually follows from the assumption that $y_{\mathrm{rad}}(\boldsymbol{x})$ is Gaussian by the Central Limit Theorem [45, 60]. This simplifies the analysis by requiring only that the mean and variance of the output test statistic be computed. A Gaussian probability distribution is then fitted to these moments and appropriate false alarm and signal detection probabilities are calculated. This approximation is asymptotically exact as the time–bandwidth product becomes arbitrarily large. Mills evaluates this approximation as a function of

---

[10]$\mathbf{I}$ is used here, and throughout, to refer to the identity matrix. The size of this matrix is given by the context, and in this case it is equal to $2TW$ [67].

Table 2.   Radiometer Output Statistics: If the output of a radiometer is modeled as a Gaussian, the parameters of its probability distribution, both the mean μ and variance σ^2, are shown below [43].

	H_0	H_1
μ	$2TW$	$2TW + 2\frac{E}{N_o}$
σ^2	$4TW$	$4TW + 8\frac{E}{N_o}$

the time–bandwidth product of the signal of interest, demonstrating its convergence for time bandwidth products larger than 10^5 [43].

To apply this technique to a burst detection problem, consider detecting a burst having energy $E = \sigma_s^2 (2TW)$, spread evenly throughout some bandwidth, W. In addition, let the covariance for the noise be $N_o\mathbf{I}$. Then, if the filter bandwidth is set to W, and the integration time to T [10], it can be shown that the output of the radiometer has the statistics shown in Table 2 [43].

If the assumption is then made that the variances at the output of the radiometer under both hypotheses are roughly equal, so that $8\frac{E}{N_o} \approx 0$, then a useful formula for the probability of detecting a particular signal results [43]. Using the definition that $Q(x)$ is the integral of the tail of a standard Gaussian probability distribution[11] from x to ∞, the probability of detection, P_D, can be expressed in terms of the false alarm rate, P_FA, as [67, 45, 43]

$$P_D = Q\left[Q^{-1}(P_{FA}) - \frac{E}{N_o\sqrt{TW}}\right]. \tag{34}$$

The biggest problem associated with the radiometer is its dependence upon knowing the noise covariance, N_o [60]. If N_o is unknown, it needs to be estimated in order to set the detection threshold to achieve a particular probability of false

---

[11]Following [58], $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} dx$

alarm [45]. Errors in this estimate will either result in a detector whose false alarm rate is too high, or whose probability of detection is too low. The solution, according to Sonnenschein and Fishman, is an orthogonal measurement not possible via the standard radiometer formulation.

> In particular, if a measurement orthogonal to the signal can be used to estimate the actual noise present during the detection interval, the noise could be tracked through all of its fluctuations. The performance of detectors, whether or not they are of the radiometric type, would be improved by this scheme since it would effectively remove the noise-level uncertainty. [60, p. 367]

Yet, having said this, Sonnenschein and Fishman present no method for making such an orthogonal measurement.

Further improvements may be attained by using knowledge of the signal structure. In the particular problem of interest to this research, the modulation type is assumed known. It should be possible to exploit this knowledge to improve the radiometer—especially since very few signaling types truly use their whole spectral band equally.

When the PSD associated with the modulation type of interest is known, a *Zero Cycle Detector*, also known as a *Spectral Matching Detector* [59], may be used [17]. This detector is the locally optimal detector for wide–sense stationary signals of near zero power. The output of this detector, expressed in terms of the PSD of the signal, $S_s(f)$, is realized by

$$y_{\text{SMD}}(\boldsymbol{x}) \;=\; \sum_{i=0}^{N-1} |X(f_i)|^2 \, S_s(f_i). \tag{35}$$

The performance of this spectral matching detector exceeds that of a radiometer [59], demonstrating the benefit of using signal information. In form and structure, this detector is implemented in the same way as the Radiometer in Figure 4—the only thing that has changed is the definition of the filter $H(f)$ on the front end. In the

case of the spectral matching detector, the filter on the front end is shaped by the square root of the PSD of the signal, $H(f) = \sqrt{S_s(f)}$.

Rostaing et al. extends this method to colored noise sequences in a paper on cyclostationary detection in colored Gaussian Noise [51]. According to Rostaing, the optimal prefilter in a colored noise environment includes an overwhitener as well, $H(f) = \frac{\sqrt{S_s(f)}}{S_n(f)}$, followed by detection as before. The use of the overwhitener causes this detector to outperform both the spectral matching detector and the radiometer in colored noise environments.

The problem with all of these improved detectors remains the fundamental problem of the radiometer: these detectors will detect any and every burst signal that passes through the filter with enough energy. While good filtering adds signal selectivity by removing those signals that are outside of the frequency band of interest, this is not enough discrimination for large interference environments where interference bursts may occupy the same band as the signal.

2.4.2 Cyclostationary Detection Techniques. If signal selectivity is desired, cyclostationary detection techniques promise it. These techniques center around $S_s^\alpha(f)$, which is assumed to be unique to the signal of interest when the cycle frequency, α, is not zero. Mathematically, this spectral correlation function will only be non–zero if a signal of interest is present—noise does not contribute at all. Thus, these techniques measure the strength of the spectral correlation for all received signals. If the correlation is present, the signal is declared to be present. If not, noise alone is declared to be present. Since neither noise nor interference possess this property, it can easily be used to discriminate between bursts of interest and bursts that are not. (See Table 1, page 18.)

From this basis, Gardner developed a locally optimum detector based on the weak signal approximation [15]. Applied to the task of detecting a cyclostationary

37

signal in white noise of variance N_o, this locally optimum detector has the form [17],

$$y_{\text{MCD}}(\boldsymbol{x}) \;=\; \frac{1}{N_o^2} \sum_{\alpha} \int_{-\infty}^{\infty} S_s^{\alpha}(f)^* \left[X_T^* \left(f - \frac{\alpha}{2} \right) X_T \left(f + \frac{\alpha}{2} \right) \right] df. \tag{36}$$

Gardner defines this to be the *multicycle cyclostationary detector*, since the sum is taken over all values of α (or cycles) where $S_s^{\alpha}(f)$ is non–zero [17]. If large integration times are available, a suboptimal single cycle detector can be created from the separate terms in this summation. This detector,

$$y_{\text{SC}}^{\alpha}(\boldsymbol{x}) \;=\; \frac{1}{N_o^2} \int_{-\infty}^{\infty} S_s^{\alpha}(f)^* \left[X_T^* \left(f - \frac{\alpha}{2} \right) X_T \left(f + \frac{\alpha}{2} \right) \right] df \tag{37}$$

evaluates Eqn. (36) at only one particular value of α, such as twice the carrier or the symbol rate [17]. When $\alpha = 0$, this single–cycle detector is identical to the spectral matching detector presented in the last section. As a result, the zero cycle frequency, as $\alpha = 0$ is termed, is highly susceptible to interference. The real power of this detector lies in those terms where $\alpha \neq 0$. In such cases, the expected response of this detector is zero for all interfering signals that do not have the cycle frequency α.

Gardner also notes, in passing, the detector that would be appropriate if the noise plus interference were not white [17]. This detector has the form [17],

$$y_{\text{MCD-CN}}(\boldsymbol{x}) \;=\; \sum_{\alpha} \int_{-\infty}^{\infty} \frac{S_s^{\alpha}(f)^*}{S_n \left(f - \frac{\alpha}{2} \right) S_n \left(f + \frac{\alpha}{2} \right)} \left[X_T^* \left(f - \frac{\alpha}{2} \right) X_T \left(f + \frac{\alpha}{2} \right) \right] df. \tag{38}$$

Rostaing et al. discusses this detector in more detail in [51].

Cyclostationary detectors are often superior in performance to radiometric detectors because they are signal specific [17]. Thus, in the presence of interfering signals, this detector tends to respond only when a signal is present having the cycle frequency of interest. For signals with different carriers or symbol rates from their

interferers, this improvement can be quite substantial. As Gardner puts it, these detectors,

> ...have many advantages over radiometry, including the ability to perform signal timing measurement, discriminate against signals not of interest using sufficiently long collects, and reduce sensitivity to unknown and changing background noise level and interference activity. [17, p. 898]

The primary problem with these techniques, as noted in the quote above, is that they require "sufficiently long collects." In the case of burst signal detection, it may not be possible to obtain "sufficiently long collects." In addition, these detectors only "reduce" the "sensitivity to unknown and changing background noise level and interference activity." As this dissertation will show, other techniques can be used to eliminate this sensitivity entirely by "making a measurement orthogonal to the signal ..." [60, p. 367].

2.4.3 A Better Detector. The ideal detector depends upon both the signal environment and the goals of the designer.

If the desire is to detect burst signals in white noise, energy detectors are highly effective. Given that the typical burst signal has plenty of signal to noise ratio, these detectors tend to work quite well. Even better, their simplicity makes them easy to design and implement. The only problem with these energy detectors is their total lack of signal selectivity.

If the desire is to selectively detect a specific signal in the presence of interference, some form of feature detection unique to the signal of interest is required. Current state of the art feature detectors, however, do not provide this selectivity without "sufficiently long collects" and integration times. While these techniques are selective, they only "reduce" the impact of the changing background level. A truly selective detector would not just reduce, but would totally eliminate the impact a changing changing background level has on the false alarm rate.

Section 3.4.2 presents a detector that meets all of these criteria. It is both applicable to burst signals, and much more selective than current cyclostationary detectors in burst interference environments. Such a detector is required in high interference environments where both bursts of signal and of interference are present.

2.5 Summary

In this chapter, two approaches to signal processing for digital communications were presented—a stationary approach and a cyclostationary one. In general, the stationary approach had a more thorough development, while the cyclostationary approach accounted for more of the properties in the signal. These extra properties were then used by others to produce symbol estimators for PAM signals [65], interference resistant TDOA estimators [24], and signal selective detection algorithms [17]. Yet neither approach, either stationary or cyclostationary, is truly appropriate for burst signal communications owing to the fact that burst signals have only a finite duration.

Several possibilities have been left open for such a new approach. For example, the optimal symbol estimator for a real PAM waveform made no use of the fact that such signals are typically modulated by a carrier, introducing an additional, unexploited, spectral redundancy [65, 2]. Indeed, although Berger and Tufts' work continues to be referenced today [3], the assumptions underlying it have yet to be revisited. Further, no practical TDOA estimator has been proposed for a colored noise environment that uses all of the cycle frequencies available to it.[12] In particular, the problem with the unknown $S_x^\alpha(f)$ value, used in all previous cyclic TDOA estimators, gets worse in a burst situation where observation time is not available to estimate this function. Finally, classical cyclostationary presence detectors, which are signal selective for lengthy signals, lose their selectivity for short signals where

---

[12]Sec. 4.3.2 will demonstrate the failures of other cyclic TDOA estimators in colored noise.

integration is not possible.[13] Each of these problems leaves open the potential of a new approach to burst communication signal processing that removes the limitations associated with longer duration signals.

The next chapter, therefore, will address these problems by returning to first principles and developing a linear subspace approach to each of these problems. This approach will lead to improvements in every one of these application areas.

---

[13]Sec. 4.4 demonstrates this.

III. Theoretical Development

Since this research focuses on extending cyclostationary theory to time–limited signals, a proper theoretical development cannot start where other cyclostationary developments have started—with the definition of a cyclostationary random process. This definition assumes an infinitely long signal, and burst signals are not quite that long. Instead, a finite length signal will be assumed from the beginning. Linear subspace theory, which is much more appropriate to describing burst signals of a finite length, will be used instead to both describe the signal and to develop algorithms appropriate for it.

This chapter will first present a new way of analyzing the received signal. Then, the remaining discussion will be built upon this analysis. The first step, then, will be to determine the unknown parameters. Estimates of these unknowns will then lead to both optimal filters and a maximum likelihood TDOA estimation algorithms. Finally, hypotheses tests will derived from this development to test for the presence of a signal. Each of these three application areas will build upon the received signal analysis derived in the first section, demonstrating a new approach to burst communication signal processing.

3.1   Spectral Subspace Theory

The first step to developing a linear subspace approach to burst signal processing is to describe the subspace that a PAM signal fits within. The fact that a subspace approach is possible is easily seen by considering the basis vectors,

$$\boldsymbol{v}_{ki} \;\; = \;\; \psi\left(t - kT_s - \tau\right)\cos\left(2\pi f_c t + \theta\right) \tag{39}$$

$$\text{and } \boldsymbol{v}_{kq} \;\; = \;\; \psi\left(t - kT_s - \tau\right)\sin\left(2\pi f_c t + \theta\right), \tag{40}$$

corresponding to the in–phase and quadrature components of the signal respectively. These basis vectors span a vector subspace containing every possible signal. The problem with this subspace lies in the fact that these basis vectors depend upon the unknown delay, τ, and carrier phase, θ, making them difficult to use in practice. Therefore, this section presents an appropriate basis for the signal in frequency together with a probability distribution function which can be used to describe the received signal in terms of the signal of interest, the interference, and the noise.

Demonstrating the subspace that a signal resides within, in frequency, means that the signal model presented in Eqn. (1) on page 7,

$$
\begin{aligned}
x\left(t\right) \;=\; n\left(t\right) \;+\; & A \sum_{n=0}^{N_s-1} \Re\left\{d_n\right\} \psi\left(t - nT_s - \tau\right) \cos\left(2\pi f_c t + \theta\right) \\
- \; & A \sum_{n=0}^{N_s-1} \Im\left\{d_n\right\} \psi\left(t - nT_s - \tau\right) \sin\left(2\pi f_c t + \theta\right),
\end{aligned}
\tag{41}
$$

must first be transformed into frequency by the Fourier transform. Since burst communications are of interest, the signal may be assumed to be time limited and therefore must fall within some observation window, $t \in \left(-\frac{T}{2}, \frac{T}{2}\right)$. In such cases, the infinite time Fourier transform of the signal is identical to the time–limited Fourier transform of the signal. However, the time–limited transform is still required because the noise term, $n\left(t\right)$, may or may not be time limited. The time–limited Fourier transform of this signal shall be defined as before, in Eqn. (4), by the integral,

$$
X_T\left(f\right) \;\triangleq\; \int_{-\frac{T}{2}}^{\frac{T}{2}} x\left(t\right) e^{-j2\pi f t} dt.
\tag{42}
$$

If we define $N_T(f)$ to be the time–limited Fourier transform of the noise plus interference term, and exploit the linearity of the Fourier transform we get the following:

$$
\begin{aligned}
X_T(f) &= N_T(f) + \frac{A}{2} \sum_{n=0}^{N_s-1} \Re\{d_n\} \int_{-\infty}^{\infty} \psi(t - nT_s - \tau) \left[e^{j2\pi f_c t + j\theta - j2\pi ft} + e^{-j2\pi f_c - j\theta - j2\pi ft} \right] dt \\
&\quad + j\frac{A}{2} \sum_{n=0}^{N_s-1} \Im\{d_n\} \int_{-\infty}^{\infty} \psi(t - nT_s - \tau) \left[e^{j2\pi f_c t + j\theta - j2\pi ft} - e^{-j2\pi f_c - j\theta - j2\pi ft} \right] dt.
\end{aligned}
$$

Letting $u = t - nT_s - \tau$, $du = dt$

$$
\begin{aligned}
&= N_T(f) + \frac{A}{2} e^{j\theta} \sum_{n=0}^{N_s-1} e^{-j2\pi(f-f_c)\tau} \left[d_n e^{-j2\pi(f-f_c)nT_s} \right] \int_{-\infty}^{\infty} \psi(u) e^{-j2\pi(f-f_c)u} du \\
&\quad + \frac{A}{2} e^{-j\theta} \sum_{n=0}^{N_s-1} e^{-j2\pi(f+f_c)\tau} \left[d_n^* e^{-j2\pi(f+f_c)nT_s} \right] \int_{-\infty}^{\infty} \psi(u) e^{-j2\pi(f+f_c)u} du
\end{aligned}
$$

At this point it makes sense to simplify this transform by defining some terms. First, let $\Psi(f)$ be the Fourier transform of the pulse shape,[1]

$$
\Psi(f) \triangleq \int_{-\infty}^{\infty} \psi(t) e^{-j2\pi ft} dt. \tag{43}
$$

Second, define $D(z)$ to be the z–Transform of the transmitted data symbols,

$$
D(z) \triangleq \sum_{n=0}^{N_s-1} d_n z^{-n}. \tag{44}
$$

Now, if we substitute these definitions into our development, they simplify the expression for $X_T(f)$,

$$
\begin{aligned}
X_T(f) &= N_T(f) + \frac{Ae^{j\theta}}{2} e^{-j2\pi(f-f_c)\tau} \Psi(f - f_c) D\left(e^{j2\pi(f-f_c)T_s} \right) \\
&\quad + \frac{Ae^{-j\theta}}{2} e^{-j2\pi(f+f_c)\tau} \Psi(f + f_c) D^*\left(e^{-j2\pi(f+f_c)T_s} \right).
\end{aligned}
$$

---

[1]See Appendix A for a discussion of some common pulse shapes and their Fourier transforms.

In practical applications $\Psi(f)$ is band-limited and so only the first term contributes for positive frequencies.[2] Thus when $f > 0$,

$$X_T(f) = \frac{Ae^{j\theta}}{2} e^{-j2\pi(f-f_c)\tau} \Psi(f - f_c) D\left(e^{j2\pi(f-f_c)T_s}\right) + N_T(f). \tag{45}$$

Two redundancies can be seen from this equation. This first results from the fact that $D\left(e^{j2\pi(f-f_c)T_s}\right)$ is periodic, and the second from the fact that it is conjugate symmetric. To see this, let $X_s(f)$ refer to that portion of the received signal caused by signal alone, $X_s(f) \triangleq X_T(f) - N_T(f)$, ignoring the noise for a moment.

The first redundancy is found between any two positive frequencies, f and $f + \frac{k}{T_s}$ for integers k. For two such frequencies,

$$\begin{bmatrix} X_s(f) \\ X_s\left(f + \frac{k}{T_s}\right) \end{bmatrix} = A \frac{e^{j\theta} e^{-j2\pi(f-f_c)\tau}}{2} \begin{bmatrix} \Psi(f - f_c) \\ e^{-j2\pi\frac{k\tau}{T_s}} \Psi\left(f + \frac{k}{T_s} - f_c\right) \end{bmatrix} D\left(e^{j2\pi(f-f_c)T_s}\right),$$

$$\tag{46}$$

and the Fourier transform of the data within the signal is completely redundant. This is caused by the fact that the signal was created by linearly modulating an impulse stream and thus $D\left(e^{j2\pi(f-f_c)T_s}\right)$ is a periodic function of f. This phenomenon was first noticed by Nyquist [46], and subsequently exploited by Berger and Tufts in their filter development [2, 65]. The implication, presented by Nyquist, is that any bandwidth larger than $\frac{1}{T_s}$ contains redundancies [46]. Therefore, this is commonly known as the *Nyquist Minimum Bandwidth* of a complex signal.

The second redundancy can be seen about the carrier frequency when d_n is a real–valued sequence. Using the fact that the z–transform of a real sequence, and similarly the Fourier transform of the real signal $\psi(t)$, are both conjugate symmetric

---

[2]Fig. 60 in Appendix A shows that several common pulse–shapes are limited to a bandwidth on the order of $\frac{2}{T_s}$. That means that as long as $f_c \gg \frac{1}{T_s}$ this approximation holds.

about zero [63], we have

$$
\begin{bmatrix} X_s\left(f\right) \\ X_s^*\left(2f_c - f\right) \end{bmatrix} = A\frac{e^{j\theta}e^{-j2\pi(f-f_c)\tau}}{2}\begin{bmatrix} \Psi\left(f - f_c\right) \\ e^{-j2\theta}\Psi\left(f - f_c\right) \end{bmatrix} D\left(e^{j2\pi(f-f_c)T_s}\right). \quad (47)
$$

This redundancy was also noticed by Nyquist, yet not exploited by Tufts. When combined with Eqn. (46), this equation implies that the minimum bandwidth of an underlying communications signal having real symbols is $\frac{1}{2T_s}$ [46].

From these two equations, Eqns. (46) and (47), a subspace can be described in frequency within which the underlying signal must lie. This subspace includes a periodic redundancy and, for real–valued signals, a redundancy about the carrier as well. This is best illustrated by two examples, one for BPSK signaling and a second one for QPSK signaling.

Example 1 (Redundancies within a BPSK Signal) *Consider a signal constrained to lie within a baseband bandwidth of $W \leq \frac{1}{T_s}$, and having symbols in the BPSK symbol set $\mathcal{D} = \{\pm 1\}$. In this case the spectral redundancies associated with the frequency f can be written as*

$$
\begin{bmatrix} X_s\left(f\right) \\ X_s\left(f + \frac{1}{T_s}\right) \\ X_s^*\left(2f_c - f\right) \\ X_s^*\left(2f_c - f - \frac{1}{T_s}\right) \end{bmatrix} = A\frac{e^{j\theta}e^{-j2\pi(f-f_c)\tau}}{2}
\tag{48}
$$

$$
\times \begin{bmatrix} \Psi\left(f - f_c\right) \\ e^{-j2\pi\frac{\tau}{T_s}}\Psi\left(f + \frac{1}{T_s} - f_c\right) \\ e^{-j2\theta}\Psi\left(f - f_c\right) \\ e^{-j2\theta}e^{-j2\pi\frac{\tau}{T_s}}\Psi\left(f + \frac{1}{T_s} - f_c\right) \end{bmatrix} D\left(e^{j2\pi(f-f_c)T_s}\right)
$$

whenever f is constrained to lie within the first half of the bandwidth of this signal,
$f_c - \frac{1}{T_s} \le f \le f_c$.

Remember that $X_s(f)$ only refers to that component of the received signal due to signal alone. The actual received vector includes appropriate noise terms as well, even though those terms have been dropped for the moment.

Example 2 (Redundancies within a QPSK Signal) *Quadrature Phase Shift Keyed (QPSK) signals are defined as those having a symbol set $\mathcal{D} = \{\pm 1, \pm j\}$. If the pulse shape, $\Psi(f)$, is bandlimited to a width of $W \le \frac{1}{T_s}$, then the redundancy present in this system is seen by*

$$
\begin{bmatrix} X_s(f) \\ X_s\left(f + \frac{1}{T_s}\right) \end{bmatrix} = \frac{A}{2} e^{j\theta} e^{-j2\pi(f-f_c)\tau} \begin{bmatrix} \Psi(f - f_c) \\ e^{-j2\pi\frac{\tau}{T_s}} \Psi\left(f + \frac{1}{T_s} - f_c\right) \end{bmatrix} D\left(e^{j2\pi(f-f_c)T_s}\right)
$$

(49)

for $f_c - \frac{1}{T_s} < f < f_c$.

Unlike the real baseband example where the signal existed within a one dimensional vector subspace of a four dimensional space for each frequency, Eqn. (49) describes the signal lying within a one dimensional subspace of a two dimensional vector space only.

The one thing that separates these signal representations from being true vector subspaces is the fact that θ and τ are unknown at the receiver. This was the problem with defining these subspaces in time, such as in Eqns. (39) and (40). By switching to frequency, the θ and τ terms have become constant multipliers that can be factored out of these expressions. This will allow us to use these vector subspaces in frequency, even though they are not fully specified.

The rest of this dissertation assumes that the signal can be described by either Eqn. (48) or Eqn. (49). Signals that cannot be described by one of these two

equations can often be described in a similar manner. Once an appropriate basis set is specified, many of the methods presented in this chapter will still follow for other signaling types as well.

3.1.1 Compact Representation. While the equations above describe redundancies associated with one particular frequency, many frequency components are required to describe a signal of interest. Indeed, it is readily proved that a complete basis for the signal requires a minimum of N_s elements for a real signal and $2N_s$ elements for a complex one, not just the single vectors describing an arbitrary frequency, f, presented above. This subsection, therefore, presents the transition from examining a single frequency to examining a set of frequencies. This subsection also introduces the transition to vector notation that comes with this switch.

In order to represent the whole frequency bandwidth containing the signal, we need to somehow sample the Fourier transform of the received signal across its bandwidth. But how many samples are necessary? If we define N_f to be the minimum number of complex frequency samples required to span the Nyquist minimum bandwidth of the signal, then $N_f = \frac{1}{2}N_s$ samples would be required to reconstruct $D\left(e^{j2\pi(f-f_c)T_s}\right)$ when the $\{d_n\}$ are real, and $N_f = N_s$ samples would be required when they are complex. The factor of two difference between the number of frequency samples required to span the vector space and the size of the basis comes from the fact that frequency samples are complex, whereas the $\{d_n\}$ values may or may not be complex.

The next step is to go from this unspecified basis which can represent $D\left(e^{j2\pi(f-f_c)T_s}\right)$ in frequency to one that can represent the entire signal bandwidth. From the redundancy equations presented in the last subsection, every value of $D\left(e^{j2\pi(f-f_c)T_s}\right)$ directly determines the signal component of m frequencies, where m is determined by the bandwidth of the signal and by the redundancy equations appropriate to it. For example, using the BPSK redundancy equation given in Eqn. (48),

$m = 4$ frequency samples are required for each $D\left(e^{j2\pi(f-f_c)T_s}\right)$ sample, while only $m = 2$ would be required for the QPSK example in Eqn. (49). That means that mN_f complex frequency samples are sufficient to describe the received signal across its entire frequency band. This is equivalent to describing m copies of the signal's Nyquist minimum bandwidth.

As for which mN_f frequencies need to be chosen, the simplest option is just to uniformly sample all of the frequencies in the signal at a spacing of $\frac{1}{T}$ Hz apart in frequency—corresponding to the resolution of a discrete Fourier transform. The first N_f of these frequency samples, or equivalently the first non–redundant set, shall be denoted by $f_1, f_2, \ldots, f_i, \ldots, f_{N_f}$. The rest of the $(m-1)N_f$ frequencies will be specified as linear functions of these first N_f frequencies.

Next, organize the Fourier transform of the data, $X_T(f)$, into a vector, $\boldsymbol{x}$, of length mN_f. For the BPSK example, this vector would look something like,

$$
\boldsymbol{x} \;=\; \begin{bmatrix}
X(f_1) \\
X(f_2) \\
\vdots \\
X(f_{N_f}) \\
X\left(f_1 + \frac{1}{T_s}\right) \\
\vdots \\
X^*(2f_c - f_1) \\
\vdots \\
X^*\left(2f_c - f_1 + \frac{1}{T_s}\right) \\
\vdots
\end{bmatrix}. \tag{50}
$$

Note that the second half of this vector was conjugated. This is only appropriate for BPSK systems where the real baseband redundancy, given in Eqn. (47), applies. In all other cases conjugation is not necessary. The noise vector, $\boldsymbol{n}$, is defined

identically, with the exception that it refers to the noise plus interference component of the received signal and not the received signal itself. Both of the vectors, $\boldsymbol{x}$ and $\boldsymbol{n}$, have the same dimension, $mN_f \times 1$.

To deal with the data component, organize the data $D(z)$ into a similar vector, $\boldsymbol{d}$, having only N_f dimensions, such as

$$\boldsymbol{d} \triangleq \left[\begin{array}{cccc} D\left(e^{j2\pi(f_1-f_c)T_s}\right) & D\left(e^{j2\pi(f_2-f_c)T_s}\right) & D\left(e^{j2\pi(f_3-f_c)T_s}\right) & \cdots \end{array} \right]^T. \tag{51}$$

This data component is first shaped by a pulse weight matrix formed from the $\Psi(f_i - f_c)$ terms in the redundancy equations. This shaping matrix shall be referred to as $\boldsymbol{\Psi}$. The contributions of τ and θ are then grouped into the unitary, and diagonal, matrix $\mathbf{R}_\phi$, where $\boldsymbol{\phi}$ refer to these complex phase values. (Examples of all of these matrices, for a BPSK system, can be found in Appendix B.) This leaves only the amplitude, $\frac{A}{2}$, scaling the whole expression.

Putting all of these vector and matrix quantities together, the received data vector may be written as,

$$\boldsymbol{x} = \frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} + \boldsymbol{n}. \tag{52}$$

As defined, the dimensions of these terms are,

$$\boldsymbol{x} \; : \; mN_f \times 1,$$
$$\boldsymbol{n} \; : \; mN_f \times 1,$$
$$\boldsymbol{d} \; : \; N_f \times 1,$$
$$\boldsymbol{\Psi} \; : \; mN_f \times N_f,$$
$$\text{and } \mathbf{R}_\phi \; : \; mN_f \times mN_f.$$

These vector and matrix expressions will make it simpler to manipulate the underlying redundancies in the following sections.

It should be stressed, before going any further, that $\mathbf{R}_\phi$ is not fully specified. $\mathbf{R}_\phi$ contains the unknown carrier phase angle, θ, together with the unknown symbol epoch, τ. These values will need to be estimated as part of any algorithm that uses this model.

3.1.2   The Distribution of $\boldsymbol{x}$.   In Eqn. (52) above, there are three unknown quantities, $\boldsymbol{n}$, $\boldsymbol{\phi}$, and $\boldsymbol{d}$. Proper statistical analysis will depend upon knowing whether each of these quantities is random or deterministic and, if random, what its probability distribution is.

The probability distribution of the noise plus interference term, $\boldsymbol{n}$, is constrained by the problem statement to be a multi–variate Gaussian random variable. In particular, it has zero mean and a known, (approximately) diagonal covariance matrix $\boldsymbol{R}_n$,

$$(\mathbf{R}_n)_{ii} \triangleq \mathcal{E}\left\{\left|\int_{-\frac{T}{2}}^{\frac{T}{2}} n\left(t\right) e^{-j2\pi ft} dt\right|^2\right\} \approx TS_n\left(f\right). \tag{53}$$

The approximation in Eqn. (53) introduces a bias into $\mathbf{R}_n$ that becomes significant when $S_n\left(f\right)$ changes "rapidly" [8]. Since this research assumes this matrix is known, there are two ways of dealing with this problem. The first is to estimate this covariance from the received data when the signal is absent. Such estimates are easily constructed in an unbiased manner, yet all such estimates suffer from uncertainty resulting from the estimation process. An alternative is to use the biased approximation in Eqn. (53), and to allow T to be large enough that this bias is negligible.

This latter approach will be used for the experiments in Chapt. IV to avoid the estimation uncertainty.[3]

The probability distribution of $\boldsymbol{\phi}$ is also constrained—it must be a deterministic, but unknown, parameter. Should $\boldsymbol{\phi}$ be treated as random, then $x(t)$ becomes a stationary process rather than a cyclostationary one [32,17].

This leaves the probability distribution of $\boldsymbol{d}$.

Prior to the z–transform, the probability distribution of $\{d_n\}$ is usually well specified by the modulation type of interest. In digital modulation schemes d_n is chosen from some finite set of elements, $\mathcal{D}$, where each element has some known probability. The d_n's may be entirely independent, or they may have some known correlation. They may be biased towards some values within $\mathcal{D}$, or uniformly distributed. The correlation in this data sequence can be expressed spectrally as,

$$S_d\left(e^{j2\pi f}\right) \quad \triangleq \quad \lim_{N_s\to\infty} \frac{1}{N_s} \mathcal{E}\left\{ \left| \sum_{n=0}^{N_s-1} d_n e^{-j2\pi fn} \right|^2 \right\} \tag{54}$$

In general, however, most communications systems transmit uncorrelated symbols, leaving this value constant. Either way, the probability distribution of $\{d_n\}$ is easily known from the modulation parameters.

The probability distribution of $\boldsymbol{d}$, the z–transform of $\{d_n\}$, is not so obvious. A first order approximation would be to treat this probability distribution as a multivariate Gaussian. Looking a little further, since $\boldsymbol{d}$ is constructed from a sum of random numbers, it may actually be Gaussian for reasonably large number of symbols, N_s, by the Central Limit Theorem. If $\boldsymbol{d}$ is truly Gaussian, all that is required to specify its probability distribution is its mean, $\boldsymbol{0}$ for most modulation types of interest, and its variance, $\mathbf{R}_d = N_s\mathbf{I}$ for statistically independent values of d_n having

---

[3]Appendix I, where application is made for small values of T, is the one exception to this rule. In that appendix, the diagonal is calculated in an exact manner.

unity magnitude. If the d_n are not statistically independent, an (approximately) diagonal matrix with elements,

$$(\mathbf{R}_d)_{ii} \approx N_s S_d \left(e^{j2\pi(f_i - f_c)T_s} \right), \tag{55}$$

may be used instead.

The following theorem is introduced to demonstrate that the central limit theorem does apply to $\boldsymbol{d}$ as $N_s \to \infty$, justifying the assumption that $\boldsymbol{d}$ is Gaussian.

Theorem 1 (The Data Vector is Asymptotically Gaussian) *Consider a message composed of N_s symbols, d_n, where d_n is drawn randomly from some finite set of symbols, $\mathcal{D}$, each having finite energy. Assume also that the d_n have zero mean. Then the discrete time Fourier transform of the data, examined from any angle, ϕ, and any radian frequency, $\omega = 2\pi \left(f - f_c \right) T_s$,*

$$\Re \left\{ e^{j\phi} D \left(e^{j\omega} \right) \right\} \triangleq \Re \left\{ e^{j\phi} \sum_{n=0}^{N_s-1} d_n e^{-j\omega n} \right\},$$

is asymptotically Gaussian as $N_s \to \infty$ as long as $\mathcal{E} \left\{ \Re \left\{ d_n e^{j\phi} e^{-j\omega n} \right\}^2 \right\} > 0$ for each n.

Proof: *To prove this, start by letting g_n be defined as one term of the sum composing $\Re \left\{ e^{j\phi} D \left(e^{j\omega} \right) \right\}$,*

$$g_n \triangleq \Re \left\{ d_n e^{j\phi} e^{-j\omega n} \right\}, \tag{56}$$

and let σ_n^2 equal the variance of g_n, $\sigma_n^2 \triangleq \mathcal{E} \left\{ g_n^2 \right\}$. Then the variance of the sum, defined as V_{N_s}, is

$$V_{N_s} \triangleq \mathcal{E} \left\{ \sum_{n=0}^{N_s-1} g_n^2 \right\} = \sum_{n=0}^{N_s-1} \sigma_n^2 \tag{57}$$

53

According to the Lindeberg–Feller Central Limit Theorem [38], $\sum g_n$ will be asymptotically Gaussian as $N_s \to \infty$ if and only if for all $\epsilon > 0$ there exists N such that, for all $N_s > N$,

$$\lim_{N_s \to \infty} \frac{1}{V_{N_s}^2} \sum_{n=0}^{N_s-1} \int_{|g_n| \geq \epsilon V_{N_s}} g_n^2 dF_n(g_n) = 0. \tag{58}$$

Thus, to prove Gaussian convergence, one need only find this value of N that satisfies Eqn. (58) for every $\epsilon > 0$.

Using the fact that each symbol has a finite energy, there exists a limit, G, such that $|d_n| < G$ with probability 1. Then, given this limit, the value of N that makes Eqn. (58) hold is $\frac{G}{\epsilon \min_n \sigma_n}$. To see that this is true, consider that, whenever $N_s > N$,

$$G = \epsilon N \min_n \sigma_n < \epsilon N_s \min_n \sigma_n < \epsilon V_{N_s}. \tag{59}$$

Looking at the integral, we know that

$$\int_{|g_n| \geq G} g_n^2 dF_n(g_n) = 0, \tag{60}$$

since $|g_n| < G$ with probability one. Further, since $\epsilon V_{N_s} > G$,

$$\int_{|g_n| \geq \epsilon V_{N_s}} g_n^2 dF_n(g_n) \leq \int_{|g_n| \geq G} g_n^2 dF_n(g_n) \tag{61}$$

$$= 0 \tag{62}$$

Q.E.D.

While most digital communications systems meet the assumptions of this proof, there is one important exception that needs to be mentioned here: BPSK systems. That is, when $\mathcal{D} = \{\pm 1\}$ or equivalently when $\mathcal{D} \subsetneq \mathbb{R}$, then any particular σ_n^2 can

always be made to be zero by examining $\phi = \frac{\pi}{2} + \omega n$. A corollary to this theorem, that $D\left(e^{j\omega}\right)$ is Gaussian when a limited portion of the σ_n^2 values are zero, can be made to apply in these cases.

The problem remains when σ_n^2 is zero for all n, such as when $\phi = \frac{\pi}{2}$ and $\omega = \pm n\pi$. In this particular case, the theorem and general method breaks down and $\Re\left\{e^{j\phi}D\left(1\right)\right\}$ is only asymptotically Gaussian for $\phi \neq \frac{\pi}{2}$. This latter exception results in a discontinuity in the probability distribution of $D\left(e^{j2\pi(f-f_c)T_s}\right)$ as $N_s \rightarrow \infty$ for these signals. For this one exception, we shall approximate $D\left(e^{j2\pi(f-f_c)T_s}\right)$ as a Gaussian, and ignore the discontinuity as though it were non–existent. While this is not quite accurate, applying this assumption leaves the probability distribution simple to work with. Therefore this approximation shall be applied as necessary for BPSK signals in this research.

Other than this one exception, this proof shows that $\boldsymbol{d}$ is asymptotically Gaussian for large N_s. But what about small values of N_s, such as are appropriate in burst communications? In this case, $\boldsymbol{d}$ may not be Gaussian at all. However, it shall be *approximated* as a Gaussian. Sec. 4.1 will present some measures of how good this approximation is. In particular, for a QPSK system, this approximation is shown to be reasonable for bursts as short as $N_s = 8$ symbols.

Putting all of these probability distributions together, and given that $\boldsymbol{d}$ is approximately a multivariate Gaussian random variable, having variance $\mathbf{R}_d$ and mean $\mathbf{0}$, the probability density function of a received signal can be written as,[4]

$$
\begin{aligned}
f\left(\boldsymbol{x}, \boldsymbol{d}\right) \;=\; & \left(\frac{1}{2\pi}\right)^{\frac{mN_f + N_f}{2}} \det\left|\mathbf{R}_n\right|^{-\frac{1}{2}} \det\left|\mathbf{R}_d\right|^{-\frac{1}{2}} \qquad\qquad (63) \\
& \times\; \exp\left\{ -\frac{1}{2}\left(\boldsymbol{x} - \frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right)^\dagger \mathbf{R}_n^{-1}\left(\boldsymbol{x} - \frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right) - \frac{1}{2}\boldsymbol{d}^\dagger\mathbf{R}_d^{-1}\boldsymbol{d} \right\}
\end{aligned}
$$

---

[4]The notation, $\boldsymbol{x}^\dagger$, is used throughout to refer to the conjugate transpose of a vector or matrix $\boldsymbol{x}$.

When $\boldsymbol{x}$ is either known or measured, this probability density function is referred to as the *likelihood* function [7]. To the extent that this probability distribution, or one similar, applies when a signal is present all of the algorithms presented in this chapter will follow.

Before leaving this discussion, it remains to be shown that previous cyclostationary properties are consequences of this new model. Since all of these properties are associated with the second moment signal, a look at that moment is provided here. For any particular observation length, T, the second moments are,

$$\mathcal{E}_{\boldsymbol{x},\boldsymbol{d}}\left\{\boldsymbol{x}\boldsymbol{x}^{\dagger}\right\} = \mathcal{E}_{\boldsymbol{d}}\left\{\left(\frac{A}{2}\mathbf{R}_{\phi}\boldsymbol{\Psi}\boldsymbol{d}\right)\left(\frac{A}{2}\mathbf{R}_{\phi}\boldsymbol{\Psi}\boldsymbol{d}\right)^{\dagger}\right\} + \mathcal{E}_{\boldsymbol{n}}\left\{\boldsymbol{n}\boldsymbol{n}^{\dagger}\right\} \tag{64}$$

$$= \frac{A^2}{4}\mathbf{R}_{\phi}\boldsymbol{\Psi}\mathbf{R}_{d}\boldsymbol{\Psi}^{\dagger}\mathbf{R}_{\phi}^{\dagger} + \mathbf{R}_{n}. \tag{65}$$

Normalizing these moments by the observation length, $T = N_s T_s$, and assuming uncorrelated symbols, $\mathbf{R}_d = N_s\mathbf{I}$,

$$\frac{1}{T}\mathcal{E}\left\{\boldsymbol{x}\boldsymbol{x}^{\dagger}\right\} = \frac{A^2}{4T_s}\mathbf{R}_{\phi}\boldsymbol{\Psi}\boldsymbol{\Psi}^{\dagger}\mathbf{R}_{\phi}^{\dagger} + \frac{1}{T}\mathbf{R}_{n}, \tag{66}$$

produces the familiar moments of interest. Since the signal contribution does not change as T increases towards infinity, taking the limit as $T \to \infty$ is a trivial matter. Looking at the terms following such a limit, the diagonal of the first term is the power spectral density of the signal,

$$S_s\left(f\right) = \frac{A^2}{4T_s}\left|\Psi\left(f - f_c\right)\right|^2. \tag{67}$$

Likewise, the off–diagonals correspond to the spectral correlation functions for $\alpha = \frac{1}{T_s}$,

$$S_s^{\frac{1}{T_s}}(f) \;=\; e^{-j2\pi \frac{\tau}{T_s}} \frac{A^2}{4T_s} \Psi^* \left(f - \frac{1}{2T_s} - f_c \right) \Psi \left(f + \frac{1}{2T_s} - f_c \right), \qquad (68)$$

and, for real signals, for $\alpha = 2f_c$,

$$S_s^{2f_c}(f) \;=\; e^{j2\theta} \frac{A^2}{4T_s} \left| \Psi \left(f - f_c \right) \right|^2. \qquad (69)$$

These are the properties that have been exploited previously to produce cyclostationary algorithms. That they are identical to previous results can be seen from [16, 21].

Preserving the first two moments of $\boldsymbol{x}$ is one of the key requirements of this new model. All other moments are of no consequence, since all of the moments of a Gaussian are completely specified by the first two. As an added benefit, these are the same moments which drove the development of previous cyclostationary algorithms cited in Chapt. II. By preserving these moments, it may be possible to gain some insight into these previous algorithms, and perhaps even to improve upon them. The best part, however, is that these known moments are now *consequences* of this model, rather than the *definition* of it.

3.1.3 Consequences. The relevance of these subspace formulas lies in the fact that the signal of interest resides within a smaller subspace than the time–bandwidth product typically used [67, 43]. This means that, for a known $\mathbf{R}_\phi$, a projection operator can be immediately created, $\mathbf{P} = \mathbf{R}_\phi \boldsymbol{\Psi} \left(\boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1}$, which will project the received waveform into the signal's subspace [54]. Likewise, an alternate projection operator, $\mathbf{I} - \mathbf{P}$, will project the received waveform onto a noise only subspace.

As applied to symbol estimation, a filter similar to this projection operator will be shown to achieve the minimum mean square error among all other filters. What's new about this filter is that it can easily be applied to any modulation type of interest, not just baseband or QPSK signals. Sec. 3.2, next, will develop this filter, and then Sec. 4.2, in the next chapter, will present its capability through several examples.

Next, the probability density function in Eqn. (63) allows the application of classical statistical principles while designing maximum likelihood TDOA estimation routines for burst communication signals. Previous attempts to do this have been unsuccessful, as Gardner highlights,

> Although classical statistical principles, such as maximum likelihood, provide an alternative approach, we have not yet found such approaches to be tractable for the non–Gaussian nonstationary models of interest. [23, p. 1177]

Indeed, since only unbiased *maximum likelihood estimators* achieve the Cramér–Rao bound in estimation error, one might anticipate that a maximum likelihood estimator derived from this model would outperform all other estimators [53]. Sec. 3.3 derives such an estimate, while Sec. 4.3 shows that it does indeed outperform all other leading estimators.

Finally, when applied to the detection problem, this projection operator allows the measurement of signal energy within one subspace, leaving all other received energy in an orthogonal subspace. This was the one capability energy detection lacked, which made it fail when the amount of background noise plus interference changed. The relevance of this technique increases further with the prediction of Sonnenschein and Fishman that,

> In particular, if a measurement orthogonal to the signal can be used to estimate the actual noise present during the detection interval, the noise could be tracked through all of its fluctuations. The performance of detectors, whether or not they are of the radiometric type, would be

improved by this scheme since it would effectively remove the noise-level uncertainty. [60, p. 367]

A detector using this scheme will be derived in Sec. 3.4, and likewise tested in Sec. 4.4 of the next chapter to illustrate its potential.

From these three applications alone, the power of the probability density function given by Eqn. (63) should be obvious.

3.2   Optimal Filters for Symbol Estimation

The first order of business in dealing with any probability model is to estimate the unknown parameters. Chief among the unknown parameters in a communications problem is the data vector, $\boldsymbol{d}$. While this may not necessarily be the first order of business chronologically, optimal estimates of $\boldsymbol{d}$ are required to develop TDOA estimation and presence detection algorithms. Therefore, all other application areas are dependent upon this first one.

This section, therefore, examines several estimates for $\boldsymbol{d}$. The first estimate is a single sensor estimate, presented in Sec. 3.2.1. This estimate underlies the appropriate filter for demodulating a communications signal, and so it leads to optimal demodulation filters. The communications symbol estimation problem is then examined in detail for BPSK signals, since the form of the filter differs from Berger and Tufts' work [2]. From this examination, Sec. 3.2.2 presents a new result showing how to predict the consequences of either a poor receiving filter or poor carrier and bit synchronization. After predicting performance, theoretical bounds for estimating $\boldsymbol{d}$ are then derived in Sec. 3.2.3. Finally, Sec. 3.2.4 shows how to extend this single sensor estimate for $\boldsymbol{d}$ to multi–sensor estimates.

3.2.1   Single Sensor MMSE Filters.
There are two ways to derive estimates for $\boldsymbol{d}$, each having a different purpose in this work. The first is to derive a maximum likelihood estimate. This is the estimate of $\boldsymbol{d}$ that maximizes the likelihood function,

or equivalently that maximizes the probability density function once the random data, $\boldsymbol{x}$, has been measured. This maximum likelihood estimate is later used in TDOA estimation and presence detection to derive optimal algorithms based upon maximum likelihood principles. The second method of deriving an estimate for $\boldsymbol{d}$ is to derive a Minimum Mean Square Error (MMSE) estimate. This latter estimate is provably optimal in a Mean Square Error (MSE) sense, and is therefore useful in developing *optimal* filters. As this section will show, these two estimates are identical.

To derive the maximum likelihood estimate, let's start with the logarithm of the likelihood function, $\mathcal{L}$, and consider everything but $\boldsymbol{d}$ to be known. Because the logarithm function is monotonic, maximizing the likelihood is equivalent to maximizing its log, given by

$$
\begin{aligned}
\mathcal{L} \quad \triangleq \quad & \ln f\left(\boldsymbol{x}, \boldsymbol{d}\right) \\
= \quad & -\frac{N_f + m N_f}{2} \ln\left(2\pi\right) - \frac{1}{2}\ln\det\mathbf{R}_d - \frac{1}{2}\ln\det\mathbf{R}_n \\
& -\frac{1}{2}\left(\boldsymbol{x} - \frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right)^{\dagger}\mathbf{R}_n^{-1}\left(\boldsymbol{x} - \frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right) - \frac{1}{2}\boldsymbol{d}^{\dagger}\mathbf{R}_d^{-1}\boldsymbol{d}.
\end{aligned}
\tag{70}
$$

Next, take the gradient of this likelihood function with respect to the conjugate of the unknown vector $\boldsymbol{d}^{\dagger}$ as described in [64],[5]

$$
\nabla_{\boldsymbol{d}^{\dagger}}\mathcal{L} \quad = \quad \frac{A}{2}\boldsymbol{\Psi}^{\dagger}\mathbf{R}_\phi^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{x} - \left(\frac{A^2}{4}\boldsymbol{\Psi}^{\dagger}\mathbf{R}_\phi^{\dagger}\mathbf{R}_n^{-1}\mathbf{R}_\phi\boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)\boldsymbol{d}.
\tag{71}
$$

Using the fact that $\mathbf{R}_\phi$ and $\mathbf{R}_n$ are diagonal, and that $\mathbf{R}_\phi$ is unitary, the $\mathbf{R}_\phi$ terms inside the parentheses cancel each other out. Once simplified, we set the gradient to

---

[5]Since there has been some question as to the validity of the methods presented in [64], this solution is validated in App. E via more conventional techniques.

0 and solve for $\boldsymbol{d}$,

$$\hat{\boldsymbol{d}}_{\text{MLE}} \;=\; \frac{A}{2} \underbrace{\left(\frac{A^2}{4} \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1}}_{\text{Equalizer}} \overbrace{\boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1}}^{\text{Matched Filter}} \; \boldsymbol{x}. \tag{72}$$

This produces the maximum likelihood estimate of $\boldsymbol{d}$, $\hat{\boldsymbol{d}}_{\text{MLE}}$.

Before proceeding to the MMSE estimate, notice that this ML estimate can be broken into two pieces. The first piece, $\boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1}$, maps the received data, $\boldsymbol{x}$, onto a vector space having N_f dimensions. As further investigation will show, this part is equivalent to applying a matched filter for colored noise followed by a downconverter and sampler. Using this portion of the estimator alone results in symbol estimates corrupted by some amount of Intersymbol Interference (ISI). The second piece of this estimate, $\left(\frac{A^2}{4} \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1}$, operates on a vector space the size of the signal. Equivalently, it operates on the symbol estimates themselves—just as a Tapped Delay Line (TDL) equalizer does. Indeed, this portion of the estimate will be shown to produce a TDL equalizer, having one tap per symbol estimate. Following this portion, the resulting symbol estimates would have as much ISI removed as the signal strength allows.

Proceeding, the second estimator of interest is the minimum mean square error estimate. That is we desire $\hat{\boldsymbol{d}}_{\text{MMSE}}$ such that

$$\hat{\boldsymbol{d}}\left(\boldsymbol{x}\right)_{\text{MMSE}} \;\triangleq\; \arg\min_{\hat{\boldsymbol{d}}(\boldsymbol{x})} \mathcal{E} \left\{ \left(\hat{\boldsymbol{d}}\left(\boldsymbol{x}\right) - \boldsymbol{d} \right)^\dagger \left(\hat{\boldsymbol{d}}\left(\boldsymbol{x}\right) - \boldsymbol{d} \right) \right\} \tag{73}$$

The solution to this minimization problem is the conditional expectation, or $\hat{\boldsymbol{d}}\left(\boldsymbol{x}\right)_{\text{MMSE}} = \mathcal{E}\left\{\boldsymbol{d}|\boldsymbol{x}\right\}$, which is known for minimizing the mean square error when an *a–priori* probability distribution for $\boldsymbol{d}$ is known [53, p. 286]. Since $\boldsymbol{d}$ and $\boldsymbol{x}$ are Gaussian, the conditional probability distribution of $\boldsymbol{d}$ given $\boldsymbol{x}$ is also Gaussian

having mean,

$$\mathcal{E}\left\{\boldsymbol{d}\middle|\boldsymbol{x}\right\} = \frac{A}{2}\left(\frac{A^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)^{-1}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\boldsymbol{x}, \tag{74}$$

and variance,

$$\mathcal{E}\left\{\boldsymbol{dd}^\dagger\middle|\boldsymbol{x}\right\} = \left(\frac{A^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)^{-1}. \tag{75}$$

Therefore, the MMSE estimate of $\boldsymbol{d}$ is identical to the maximum likelihood estimate,

$$\hat{\boldsymbol{d}}\left(\boldsymbol{x}\right)_{\text{MMSE}} = \frac{A}{2}\left(\frac{A^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)^{-1}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\boldsymbol{x}. \tag{76}$$

What may not be so obvious is that this minimum mean square error estimate for $\boldsymbol{d}$ defines an optimal filter for recovering $\boldsymbol{d}$. To see this, the operation of the downconverter and sampler, which maps the $mN_f \times 1$ vector onto an $N_f \times 1$ vector, needs to be separated from the operation of the filter which applies a scale factor to each input value. It is this latter operation that we are interested in. These two operations may be separated by rewriting Eqn. (76) in terms of the individual elements composing the data estimate, $\hat{\boldsymbol{d}}\left(\boldsymbol{x}\right)_{\text{MMSE}}$.

For example, if $\{d_n\}$ is a complex valued sequence, then only the symbol rate redundancy given in Eqn. (46) applies. From this redundancy, the matrices $\mathbf{R}_\phi$, $\boldsymbol{\Psi}$ and $\mathbf{R}_n$ are defined. Using these matrices, then, we solve for $\hat{D}\left(e^{j2\pi(f-f_c)T_s}\right)$ in terms of the operation applied to each individual input frequency component of $X_T\left(f_i\right)$,

$$\begin{aligned}\hat{D}\left(e^{j2\pi(f_i-f_c)T_s}\right) &= & e^{-j\theta}e^{j2\pi(f_i-f_c)\tau} &\cdot& H\left(f_i\right)X_T\left(f_i\right) \tag{77}\\ &+& e^{-j\theta}e^{j2\pi(f_i-f_c)\tau} &\cdot& e^{j2\pi\frac{\tau}{T_s}}H\left(f_i+\frac{1}{T_s}\right)X_T\left(f_i+\frac{1}{T_s}\right).\end{aligned}$$

While the complex constants in Eqn. (77) refer to the downconversion and sampling process, the operator, $H(f)$, applied to the input waveform is the filter of interest. Using Eqn. (76) to define $H(f)$, and solving, we get the optimal filter for a complex baseband signal,

$$
H(f) = \left(\cfrac{1}{1 + \cfrac{A^2}{4T_s} \displaystyle\sum_{n=-\infty}^{\lfloor fT_s \rfloor} \cfrac{\left| \Psi\left(f - \frac{n}{T_s} - f_c\right) \right|^2}{S_n\left(f - \frac{n}{T_s}\right)}} \right) \frac{A\Psi^*(f - f_c)}{2T_sS_n(f)}.
\tag{78}
$$

When the signal is bandlimited, this filter is equivalent to Berger and Tufts' filter in Eqn. (16). Since all practical signals are bandlimited, this filter is equivalent for all practical purposes. However, this filter is only optimal for systems of complex symbols, where there are no redundancies other than the symbol rate. It is not the optimal filter for a BPSK system.

For a bandlimited BPSK system the optimal filter is still given in Eqn. 76, only the forms of $\mathbf{R}_\phi$ and $\boldsymbol{\Psi}$ have changed. These matrices, $\mathbf{R}_n$, $\mathbf{R}_\phi$, $\boldsymbol{\Psi}$, and $\mathbf{R}_d$ are given by Eqn. (48) earlier, and samples of them are provided in App. B. To calculate the form of this filter, we again arrange Eqn. (76) in terms of the individual components composing one data frequency estimate,

$$
\begin{aligned}
\hat{D}\left(e^{j2\pi(f_i - f_c)T_s}\right) =\ & e^{-j\theta}e^{j2\pi(f_i - f_c)\tau} \cdot H(f_i)\,X_T(f_i) \\
& + e^{-j\theta}e^{j2\pi(f_i - f_c)\tau} \cdot e^{j2\pi\frac{\tau}{T_s}} H\left(f_i + \tfrac{1}{T_s}\right) X_T\left(f_i + \tfrac{1}{T_s}\right) \\
& + e^{j\theta}e^{j2\pi(f_i - f_c)\tau} \cdot H^*(2f_c - f_i)\,X_T^*(2f_c - f_i) \\
& + e^{j\theta}e^{j2\pi(f_i - f_c)\tau} \cdot e^{j2\pi\frac{\tau}{T_s}} H^*\left(2f_c - f_i - \tfrac{1}{T_s}\right) X_T^*\left(2f_c - f_i - \tfrac{1}{T_s}\right).
\end{aligned}
\tag{79}
$$
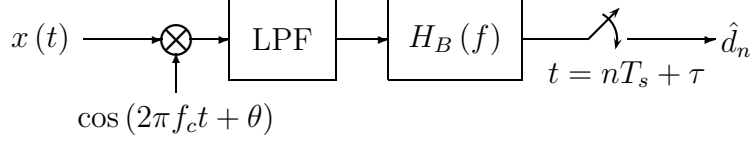
Figure 5.   Baseband MMSE (Berger's) System Diagram

Then, expanding Eqn. (76) to solve for $H(f)$, we see that the minimum mean square error filter for real digital modulations is

$$H_{\text{MMSE}}(f) = \frac{\frac{A}{2}\frac{\Psi^*(f-f_c)}{T_s S_n(f)}}{1 + \frac{A^2}{4}\frac{|\Psi(f-f_c)|^2}{T_s S_n(f)} + \frac{A^2}{4}\frac{\left|\Psi\left(f+\frac{1}{T_s}-f_c\right)\right|^2}{T_s S_n\left(f+\frac{1}{T_s}\right)} + \frac{A^2}{4}\frac{|\Psi(f_c-f)|^2}{T_s S_n(2f_c-f)} + \frac{A^2}{4}\frac{\left|\Psi\left(f_c-f-\frac{1}{T_s}\right)\right|^2}{T_s S_n\left(2f_c-f-\frac{1}{T_s}\right)}},$$

(80)

for $f \in \left(f_c - \frac{1}{T_s}, f_c\right)$.

Both similarities and differences exist between this filter and Berger and Tufts' MMSE filter in Eqn. (16). Both of these filters meet the structure required by an optimal receiver filter—they each factor into a matched filter followed by an equalizer that is periodic in frequency [12]. Each of the equalizers can be implemented by a TDL. The matched filter in each case is identical to the one presented in Sec. 2.2. The difference between these two filters lies in the equalizer. The MMSE filter for BPSK signals has an equalizer that is symmetric about the carrier frequency. This means it can be implemented with a real valued TDL equalizer, while Berger and Tufts' equalizer is not necessarily real valued.

The difference between these two filters can be seen in the design of a system that would implement them. Fig. 5 shows a block diagram of a system implementing Berger's MMSE filter. In this diagram, the signal is first multiplied by a cosine matching its carrier frequency and phase, followed by a low–pass filter (LPF). Together, these two components *downconvert* the signal and remove its carrier frequency. Following this downconverter, Berger's filter is then applied prior to the
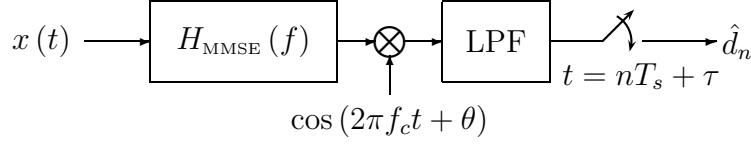
Figure 6.   Bandpass MMSE System Diagram

sampler. The alternative, shown in Fig. 6, applies the filter prior to downconverting the signal. As Sec. 4.2.1 will demonstrate, this arrangement allows the filter to mitigate narrowband interference before the downconverter makes that interference worse.

If the MMSE filter specified in Eqn. (80) truly has the *minimum* mean square error property, it should be possible to compare its results to other filters that are known to achieve the minimum mean square error, such as a true matched filter followed by a linear, adaptive, decision–directed feedback system [25]. Such adaptive equalizers are driven by the MSE performance criterion and have been shown to achieve the minimum MSE [30]. If the MMSE filter derived above and described in Eqn. (76) truly has the minimum mean square error then an adaptive algorithm will converge to the TDL portion of Eqn. (76). In the next chapter, Sec. 4.2 will demonstrate that the form of the filter above is identical to a matched filter followed by a linear adaptive equalizer.

Berger and Tufts' development did not stop when they specified the filter, they continued their development by determining the mean square error that one might measure at the output of the filter between the estimated symbol and the true symbol. Since the variance in $\hat{\boldsymbol{d}}_{\text{MMSE}}$ is given by Eqn. (75), the inverse z–transform of the variance of $\boldsymbol{d}$ yields the variance of d_n [2, 57]. Thus, in the case of a BPSK

signal, the mean square error at the output of the BPSK MMSE filter should be,

$$\xi^2 \triangleq \mathcal{E}\left\{\left|d_n - \hat{d}_n\right|^2\right\} = 2T_s \int_0^{\frac{1}{2T_s}} \cfrac{1}{\begin{bmatrix} 1 & + & \frac{A^2}{4}\frac{|\Psi(f)|^2}{T_s S_n(f_c+f)} & + & \frac{A^2}{4}\frac{\left|\Psi\left(f-\frac{1}{T_s}\right)\right|^2}{T_s S_n\left(f_c+f-\frac{1}{T_s}\right)} \\ & + & \frac{A^2}{4}\frac{|\Psi(-f)|^2}{T_s S_n(f_c-f)} & + & \frac{A^2}{4}\frac{\left|\Psi\left(-f+\frac{1}{T_s}\right)\right|^2}{T_s S_n\left(f_c-f+\frac{1}{T_s}\right)} \end{bmatrix}} df.$$

(81)

Again, it is worth pointing out the assumptions that have been used so far. First, while the MMSE filter may be used for a number of signal types under a large variety of conditions, assuming the redundancy for bandlimited BPSK signals given in Eqn. (48), together with equiprobable independent bits, $\mathbf{R}_d = N_s\mathbf{I}$, led us to the conclusions in Eqn. (80) and Eqn. (81). Breaking either of these assumptions would invalidate these two equations but not the equation for the general MMSE filter given by Eqn. (76).

3.2.2 Predicting Demodulator Performance. The structure used in the previous section provides insight into how a receiver operates when it determines the underlying bit sequence. In particular we see from Eqn. (79) how each component of the signal is used to determine the underlying bits. This section will take that same equation one step further by using it to estimate the performance of an *arbitrary* filter. This performance will first be estimated in terms of mean square error, and then these estimates will be extended to bit error rates.

The approach that accomplishes this is straightforward: calculate the mean square error between $\hat{D}\left(e^{j2\pi(f-f_c)T_s}\right)$ and $D\left(e^{j2\pi(f-f_c)T_s}\right)$ as a function of frequency, f, in terms of the filter, $H(f)$, the pulse function, $\Psi(f)$ and the synchronization parameters τ and θ. Integrating this mean square error across frequency, together with some appropriate normalizations, results in an estimate of the mean square error between the input and output symbols.

Before starting, a couple of new terms need to be introduced, together with some simplifying assumptions. $H(f)$ will be used to refer to the demodulation filter, which will be allowed to be arbitrary. Next, we shall assume that the difference between the estimated symbol epoch and the true symbol epoch, also known as the synchronization error $\tau_\delta \triangleq \hat{\tau} - \tau$, is small in comparison to a symbol interval. Likewise, the difference between the estimated carrier phase and the true carrier phase, the carrier synchronization error $\theta_\delta \triangleq \hat{\theta} - \theta$, is also assumed to be small. Finally, all parameters save the data symbols are assumed known for this analysis.

Starting at the top, the fundamental quantity in this section is the mean square error. In particular, the mean square error as a function of frequency, $\xi^2\left(e^{j2\pi(f-f_c)T_s}\right)$, is defined to be,

$$\xi^2\left(e^{j2\pi(f-f_c)T_s}\right) \quad \triangleq \quad \mathcal{E}\left\{\left|\hat{D}\left(e^{j2\pi(f-f_c)T_s}\right) - D\left(e^{j2\pi(f-f_c)T_s}\right)\right|^2\right\}. \tag{82}$$

This mean square error expression can be simplified by using the fact that the signal and the noise are independent and zero mean. That means that this expression, containing signal and noise contributions, can be broken into two components, corresponding to the mean square error due to the noise plus interference process, $n(t)$, which shall be denoted $\xi_n^2\left(e^{j2\pi(f-f_c)T_s}\right)$, and the mean square error due to intersymbol interference and mis–synchronization $\xi_i^2\left(e^{j2\pi(f-f_c)T_s}\right)$,

$$\xi^2\left(e^{j2\pi(f-f_c)T_s}\right) \quad = \quad \xi_n^2\left(e^{j2\pi(f-f_c)T_s}\right) + \xi_i^2\left(e^{j2\pi(f-f_c)T_s}\right) \tag{83}$$

Having broken this expression into two components, each can be examined separately.

The first component, the error due to the noise, is given by plugging the expression for $\hat{D}\left(e^{j2\pi(f-f_c)T_s}\right)$ given by Eqn. (79) into Eqn. (82) and paying attention to the terms containing $S_n(f)$ only. Then, using the definition for $\mathbf{R}_n$ given in Eqn. (53),

this expression becomes,

$$
\begin{aligned}
\xi_n^2 \left(e^{j2\pi(f-f_c)T_s}\right) &= & \kappa^2 \left|H\left(f\right)\right|^2 TS_n\left(f\right) & \qquad (84)\\
&+& \kappa^2 \left|H\left(f+\tfrac{1}{T_s}\right)\right|^2 TS_n\left(f+\tfrac{1}{T_s}\right) &\\
&+& \kappa^2 \left|H\left(2f_c-f\right)\right|^2 TS_n\left(2f_c-f\right) &\\
&+& \kappa^2 \left|H\left(2f_c-f-\tfrac{1}{T_s}\right)\right|^2 TS_n\left(2f_c-f-\tfrac{1}{T_s}\right),&
\end{aligned}
$$

where κ has been introduced to represent an automatic gain adjustment in the demodulator necessary to minimize the mean square error in case of a filter gain mismatch. This requirement is driven by the fact that most filters, the matched filter prominent among them, are specified in a gain independent fashion. Defining κ makes it possible to compare filters with different gains by allowing the demodulator to optimally adjust the gain following the filter.

The second component, the error due to intersymbol interference and mis–synchronization is given similarly. As before, Eqn. (79) is placed into Eqn. (82), only this time the signal terms are examined instead of the noise terms. In this case, however, each of the four redundant frequencies is dependent upon the same data value, $D\left(e^{j2\pi(f-f_c)T_s}\right)$. Thus, using the expression for $\mathbf{R}_d$ given by Eqn. (55), this expression becomes,

$$
\xi_i^2\left(e^{j2\pi(f-f_c)T_s}\right) = \left|
\begin{array}{rcl}
\kappa e^{-j\theta_\delta}e^{j2\pi(f-f_c)\tau_\delta} &\cdot& AH\left(f\right)\Psi\left(f\right)\\
+\ \kappa e^{-j\theta_\delta}e^{j2\pi(f-f_c)\tau_\delta} &\cdot& Ae^{j2\pi\frac{\tau_\delta}{T_s}}H\left(f+\tfrac{1}{T_s}\right)\Psi\left(f+\tfrac{1}{T_s}\right)\\
+\ \kappa e^{j\theta_\delta}e^{j2\pi(f-f_c)\tau_\delta} &\cdot& AH^*\left(2f_c-f\right)\Psi^*\left(2f_c-f\right)\\
+\ \kappa e^{j\theta_\delta}e^{j2\pi(f-f_c)\tau_\delta} &\cdot& Ae^{j2\pi\frac{\tau_\delta}{T_s}}H^*\left(2f_c-f-\tfrac{1}{T_s}\right)\Psi^*\left(2f_c-f-\tfrac{1}{T_s}\right)\\
-\ 1 & &
\end{array}
\right|^2
$$

$$
\times \quad N_s S_d\left(e^{j2\pi(f-f_c)T_s}\right). \qquad (85)
$$

In the case of random, uncorrelated, data of unit magnitude which we've been considering,

$$S_d \left(e^{j2\pi(f-f_c)T_s} \right) = 1.$$

What we wish to know is the mean square error in an arbitrary element, d_n, of the sequence being demodulated. Eqn. (83) gives the mean square error in all of the elements as a function of frequency. Thus we divide it by the number of elements that this error is spread over, or N_s, and then take the inverse Fourier transform. The result,

$$\xi^2(\kappa) \quad = \quad 2T_s \int_0^{\frac{1}{2T_s}} \frac{\xi^2 \left(e^{j2\pi fT_s} \right)}{N_s} df, \tag{86}$$

is still a function of the unknown gain, κ.

The best performance that this filter can achieve is given by the gain, κ, that gives the smallest mean square error. To find this value of κ, we note that both Eqns. (84) and (85) are *quadratic* functions of κ. That means that there exist constants, c_1, c_2, and c_3, such that the mean square error can be written as a quadratic function of κ using these constants,

$$\xi^2(\kappa) \quad = \quad c_1\kappa^2 + c_2\kappa + c_3. \tag{87}$$

These constants can be found by first factoring the $\xi^2 \left(e^{j2\pi(f-f_c)T_s} \right)$ term in Eqn. (86) into a quadratic function of κ, and then by integrating the scalar, linear, and quadratic coefficients that result separately. The minimum of this quadratic, corresponding to the optimal MSE, is found when $\kappa = -\frac{c_2}{2c_1}$. At this point, the MSE

is,

$$\xi^2 \;=\; c_3 - \frac{c_2^2}{4c_1}. \tag{88}$$

While this is an interesting, and useful, performance measure—the bottom line in any communication system is always the bit error rate (BER) at the output of the receiver. When the mean square error is the result of a Gaussian, or approximately Gaussian, disturbance this BER can be calculated. To get there, however, the mean and variance of each symbol estimate need to be determined in order to describe this disturbance.

The mean of the symbol estimate is equivalent to the gain throughout the entire system. This *system gain* is defined by,

$$\gamma \;\triangleq\; \frac{\mathcal{E}\left\{\hat{d}_n\right\}}{d_n}, \tag{89}$$

and calculated by taking the inverse Fourier transform of $H\left(f\right)\Psi\left(f-f_c\right)$, after adjusting for phase and symbol synchronization errors,

$$\gamma \;=\; 2\kappa T_s \int_{f_c-\frac{1}{T_s}}^{f_c+\frac{1}{T_s}} \Re\left\{ e^{-j\theta_\delta} e^{j2\pi(f-f_c)\tau_\delta} H\left(f\right)\Psi\left(f-f_c\right) \right\} df. \tag{90}$$

Since the system gain is not necessarily unity, the *variance* is not necessarily the mean square error, ξ^2. Instead this variance is given by,

$$\mathcal{E}\left\{\left|\hat{d}_n - \mathcal{E}\left\{\hat{d}_n\right\}\right|^2\right\} \;=\; \mathcal{E}\left\{\left|\hat{d}_n - \gamma d_n\right|^2\right\} \;=\; \xi^2 - (1-\gamma)^2. \tag{91}$$

These two values, the mean at the output of the filter, γd_n, and the variance at the output of the filter, $\xi^2 - (1-\gamma)^2$, fully specify a Gaussian probability distribu-

tion. If the additional assumption is then applied that the symbol estimates, $\hat{d}_n$, are statistically independent, calculating the probability of bit error becomes quite straightforward. This probability is given by [58],

$$
\begin{aligned}
BER &= P\left[\hat{d}_n > 0 \mid d_n = -1\right] P\left[d_n = -1\right] + P\left[\hat{d}_n < 0 \mid d_n = 1\right] P\left[d_n = 1\right] \\
&= Q\left(\sqrt{\frac{\gamma^2}{\xi^2 - (1-\gamma)^2}}\right),
\end{aligned}
\tag{92}
$$

and is valid as long as the mean square error is caused by a Gaussian disturbance.

In general, however, ISI is not Gaussian and the probability of a bit error in ISI is much more difficult to calculate. Other techniques, such as those in [34], that use more appropriate probability distribution functions for ISI give more accurate estimates of the bit error rate under severe ISI conditions.

The strength of these formulas may not be immediately apparent, and so an illustration will help. Many modern communications systems struggle with the effects of both colored noise and multipath interference. While both can be compensated for using a MMSE filter, compensation is only possible when their contribution is known. The problem is that these contributions, multipath interference and colored noise, are difficult to estimate. Multipath interference, especially, is a well known but difficult problem in cellular communications. Further, all communications systems struggle with some amount of imperfect synchronization. What these formulas allow a designer to do is to estimate the impact of using a non–optimal solution prior to implementing and testing that solution with either expensive hardware or large quantities of computer time.

As to the validity of these formulas, simulations in Sec. 4.2.2 will compare these estimates to simulated filter performance. From those simulations, Eqn. (88) for the resulting MSE will be demonstrated to be valid under all circumstances tested. This again commends the validity of this technique.

3.2.3 Cramér–Rao Bounds. Returning to the application of a new signal model, it would be nice not only to determine how well a receiver *does* work, but also how well it *can* work given that several variables need to be estimated. The quantity used to express this theoretical limit is the Cramér–Rao bound. This bound is derived mathematically, independent of any received data, and it forms a lower limit on the mean square error achievable by any estimator [53, 7]. Demonstrating this bound will help to further demonstrate the potential of this signal model when applied to communications signals.

The Cramér–Rao bound is easily specified for a single variable, such as τ, where it is [7, 53]

$$CRB\left(\tau\right) \;=\; \frac{1}{\mathcal{E}_{\boldsymbol{x}}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau^2}\right\}}, \tag{93}$$

where $\mathcal{E}_{\boldsymbol{x}}\left\{\cdot\right\}$ refers to an expected value taken over the random variable $\boldsymbol{x}$ and $\mathcal{L}$ is the log of the likelihood function as before. This bound, however, is dependent upon the message content, $\boldsymbol{d}$. To remove this dependence, we follow D'Andrea's lead and switch to the Modified Cramér–Rao Bound (MCRB) [9]. The difference between the MCRB and the CRB is a second expectation taken over the random message variable, $\boldsymbol{d}$, as well. Thus the modified bound, for τ is,

$$MCRB\left(\tau\right) \;=\; \frac{1}{\mathcal{E}_{\boldsymbol{d}}\mathcal{E}_{\boldsymbol{x}}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau^2}\right\}}. \tag{94}$$

D'Andrea et al. prove that this bound is lower than the true CRB, and so it remains a valid lower bound on estimation error [9].

In the case where multiple parameters need to be estimated at the same time, these equations change somewhat. In this case, the CRB is a matrix quantity given

by the inverse of the matrix whose elements are,

$$(\text{FIM})_{ij} \;=\; \mathcal{E}\left\{-\frac{\partial^2}{\partial\rho_i\partial\rho_j}\mathcal{L}\right\} \tag{95}$$

where ρ_i is the ith parameter of the likelihood function. Likewise if $\boldsymbol{\rho}$ is a vector containing all of the parameters to the likelihood function, this matrix is given by,

$$(\text{FIM}) \;=\; \mathcal{E}\left\{-\nabla_{\boldsymbol{\rho}}^2\mathcal{L}\right\}. \tag{96}$$

This matrix is common to multi–sensor developments and is known as the Fisher Information Matrix (FIM) [53]. The ith diagonal element in this inverse is the Cramér–Rao bound of ρ_i. As before, taking the expectation over the random variable $\boldsymbol{x}$ produces the true CRB, whereas taking it over both $\boldsymbol{x}$ and $\boldsymbol{d}$ produces the MCRB.

Before demonstrating this technique, it is convenient to define a new $mN_f \times mN_f$ matrix, $\mathbf{F}_\delta$, to make it easier to specify the derivative of $\mathbf{R}_\phi$ with respect to τ.

$$\mathbf{F}_\delta \;\triangleq\; \frac{1}{-j2\pi}\left(\frac{\partial}{\partial\tau}\mathbf{R}_\phi\right)\mathbf{R}_\phi^\dagger \tag{97}$$

$$= \begin{bmatrix} (f_1 - f_c) & 0 & \cdots \\ 0 & (f_2 - f_c) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{98}$$

Using this matrix, $\frac{\partial}{\partial\tau}\mathbf{R}_\phi$ can be expressed as $-\frac{1}{j2\pi}\mathbf{F}_\delta\mathbf{R}_\phi$.

$\mathbf{F}_\delta$ makes it easy to express the diagonal terms of the FIM corresponding to partials with respect to τ, A, and $\boldsymbol{d}$. These diagonals are derived in App. J and

shown here as,[6]

$$\mathcal{E}_{\boldsymbol{d}}\mathcal{E}_{\boldsymbol{x}}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau^2}\right\} = A^2 \pi^2 \text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^{\dagger}\mathbf{R}_n^{-1}\mathbf{F}_{\delta}^2\right\}, \tag{99}$$

$$\mathcal{E}_{\boldsymbol{d}}\mathcal{E}_{\boldsymbol{x}}\left\{-\frac{\partial^2 \mathcal{L}}{\partial A^2}\right\} = \frac{1}{4}\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^{\dagger}\mathbf{R}_n^{-1}\right\} \tag{100}$$

$$\text{and } \mathcal{E}_{\boldsymbol{d}}\mathcal{E}_{\boldsymbol{x}}\left\{-\nabla_{\boldsymbol{d}}^2\mathcal{L}\right\}, = \frac{A^2}{4}\boldsymbol{\Psi}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}. \tag{101}$$

In addition, all of the cross terms, such as the partial with respect to A followed by τ etc., go to zero (see App. J). That makes inverting this expression to determine the Cramér–Rao bounds simple. The bounds are, therefore,

$$MCRB\left(\tau\right) = \frac{1}{A^2 2\pi^2}\frac{1}{\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^{\dagger}\mathbf{R}_n^{-1}\mathbf{F}_{\delta}^2\right\}}, \tag{102}$$

$$MCRB\left(A\right) = \frac{4}{\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^{\dagger}\mathbf{R}_n^{-1}\right\}}, \tag{103}$$

$$\text{and } MCRB\left(\boldsymbol{d}\right) = \left(\frac{A^2}{4}\boldsymbol{\Psi}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)^{-1} \tag{104}$$

Of these three expressions, the last one is the most familiar. That expression was presented earlier in Eqn. (81) to express the mean square error achieved by the BPSK MMSE filter. That means that this MMSE filter, derived under the assumption that the signal is Gaussian, achieves the theoretical limit in estimation performance.

3.2.4   Dual Sensor MMSE Filters.   The same principles that have been used in the last two sections to derive optimal single sensor filters can be extended to derive an optimal multi–sensor filter. In this subsection, we derive these multi–sensor filters, demonstrating them through the design of a two sensor MMSE filter for BPSK signals. This new development begins with adjusting the model so that it describes

---

[6]$\text{tr}\left\{\cdot\right\}$ is used here to represent the trace operator. This operator may be applied to a matrix, and returns the sum of the diagonal entries of its argument.

the multi–sensor reception problem, then the filter is derived from matrix equations as before. Finally, the subsection ends by specifying the filter in a conventional manner and stating the MSE achieved by this filter.

The first step, however, is to work with the model. Several minor changes are required to the signal model presented in Eqn. 1 to accommodate multiple sensors. The first, most obvious change, is that the signal may be present on each sensor. It may have a time delay difference, $\tau_{di} \triangleq \tau_i - \tau_0$ between sensors, a phase delay between sensors, $\theta_{di} \triangleq \theta_i - \theta_0$, and possibly a separate gain on each sensor, A_i.

$$x_i = A_i \Re \left\{ \sum_{n=0}^{N_s-1} d_n \psi \left(t - nT_s - \tau_0 - \tau_{di} \right) e^{j[2\pi f_c(t - \tau_{di}) + \theta_0 + \theta_{di}]} \right\} + n_i(t). \quad (105)$$

Assuming that all of these quantities are known, the Fourier transform of the signal on the ith sensor may be written as,

$$X_i(f) = \frac{A_i}{2} \left(e^{j\theta_{di}} e^{-j2\pi f \tau_{di}} \right) \left(e^{j\theta} e^{-j2\pi(f - f_c)\tau} \right) \Psi(f - f_c) D \left(e^{j2\pi(f - f_c)T_s} \right) + N_{T,i}(f),$$

$$(106)$$

where θ and τ have been used instead of θ_0 and τ_0 respectively.

The redundancies in this model increase linearly with the number of sensors. For a two sensor BPSK problem, for example, there are now eight received values that might provide insight into $D \left(e^{j2\pi(f - f_c)T_s} \right)$ instead of the four values we have been using. Writing these values out, however, requires first introducing a new matrix, $\mathbf{D}_i$. This matrix is very similar to $\mathbf{R}_\phi$ in that it is complex, unitary, and diagonal. The diagonal elements, however, are slightly different. For the four co–dependent

frequencies of f, this matrix is,

$$\mathbf{D}_i(f) \triangleq \begin{bmatrix} e^{j\theta_{di}}e^{-j2\pi f\tau_{di}} & 0 & 0 & 0 \\ 0 & e^{j\theta_{di}}e^{-j2\pi\left(f+\frac{1}{T_s}\right)\tau_{di}} & 0 & 0 \\ 0 & 0 & e^{-j\theta_{di}}e^{-j2\pi(2f_c-f)\tau_{di}} & 0 \\ 0 & 0 & 0 & e^{-j\theta_{di}}e^{-j2\pi\left(2f_c-f-\frac{1}{T_s}\right)} \end{bmatrix}. \tag{107}$$

As with $\mathbf{R}_\phi$ and $\mathbf{R}_n$, $\mathbf{D}_i$ is formed by replicating the matrix in Eqn. (107) until it has one row and one column for each of the mN_f frequencies that the signal occupies. Defining this matrix allows the received waveform to again be written in vector form. Only this time the form reflects multiple sensors,

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{x}_1 \end{bmatrix}}_{\boldsymbol{x}} = \frac{1}{2} \underbrace{\begin{pmatrix} \mathbf{R}_\phi & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_1\mathbf{R}_\phi \end{pmatrix}}_{\mathbf{R}_\Phi} \underbrace{\begin{bmatrix} A_0\boldsymbol{\Psi} \\ A_1\boldsymbol{\Psi} \end{bmatrix}}_{\boldsymbol{\Psi}_m} \boldsymbol{d} + \underbrace{\begin{bmatrix} \boldsymbol{n}_0 \\ \boldsymbol{n}_1 \end{bmatrix}}_{\boldsymbol{n}}. \tag{108}$$

This equation is quite similar to Eqn. (52) (page 50). Writing it in the form of Eqn. (52), however, requires four new definitions. First, define $\boldsymbol{x}$ and $\boldsymbol{n}$ to be the vectors of multi–sensor inputs and noise respectively. Then define $\mathbf{R}_\Phi$ to be the multi–sensor matrix containing all of the phase terms, including both $\mathbf{R}_\phi$ and $\mathbf{D}_1$. Finally, define $\boldsymbol{\Psi}_m$ to be a multi–sensor $\boldsymbol{\Psi}$ matrix, but this time one that includes the gain terms A_0 and A_1. Once accomplished, a received vector can be created describing the inputs from all of the sensors at once,

$$\boldsymbol{x} = \frac{1}{2}\mathbf{R}_\Phi\boldsymbol{\Psi}_m\boldsymbol{d} + \boldsymbol{n}. \tag{109}$$

A second change to the model is to admit cross correlations between the sensors. These correlations, or *cross spectral densities*, were defined in Eqn. (5) on page 13.

In order to use these densities, however, they need to be arranged in a matrix form as in,

$$\mathbf{R}_n \triangleq \begin{bmatrix} \mathbf{R}_{n,0} & \mathcal{E}\left\{ \boldsymbol{n}_0 \boldsymbol{n}_1^{\dagger} \right\} \\ \mathcal{E}\left\{ \boldsymbol{n}_1 \boldsymbol{n}_0^{\dagger} \right\} & \mathbf{R}_{n,1} \end{bmatrix}, \tag{110}$$

where the new off–diagonal elements are themselves diagonal matrices whose elements are given by

$$\mathcal{E}\left\{ \boldsymbol{n}_0 \boldsymbol{n}_1^{\dagger} \right\}_{ii} = \mathcal{E}\left\{ N_{T,0}\left(f_i \right) N_{T,1}^{*}\left(f_i \right) \right\} \approx T S_{01}\left(f_i \right). \tag{111}$$

This fundamentally changes the structure of $\mathbf{R}_n$ from a diagonal matrix to an $M \times M$ block matrix with diagonal $mN_f \times mN_f$ submatrices.

From this point the development is almost identical to the BPSK MMSE development in Sec. 3.2.1. The MMSE data estimate is derived identically to the previous development—only this time the $\mathbf{R}_{\Phi}^{\dagger} \mathbf{R}_n^{-1} \mathbf{R}_{\Phi}$ term cannot be simplified,

$$\hat{\boldsymbol{d}}_{\text{MMSE}} = \frac{1}{2} \left(\frac{1}{4} \boldsymbol{\Psi}_m^{\dagger} \mathbf{R}_{\Phi}^{\dagger} \mathbf{R}_n^{-1} \mathbf{R}_{\Phi} \boldsymbol{\Psi}_m + \mathbf{R}_d^{-1} \right)^{-1} \boldsymbol{\Psi}_m^{\dagger} \mathbf{R}_{\Phi}^{\dagger} \mathbf{R}_n^{-1} \boldsymbol{x}. \tag{112}$$

Looking at the dimensions of the components of this matrix equation, the same structure is present that was found in the single sensor filter. This time, however, the matched filter front end includes not only the downconverter and sampler, but a signal combining operation as well. The TDL equalizer has the same number of dimensions as before. This equation is factored, just like before (See Eqn. (79)), into

a set of filters operating on the received data from each sensor,

$$
\hat{D}\left(e^{j2\pi(f_i-f_c)T_s}\right)
$$

$$
\begin{aligned}
= \quad & e^{-j\theta}e^{j2\pi(f_i-f_c)\tau} \quad \cdot \quad H_0\left(f_i\right)x_0\left(f_i\right) & (113) \\
+ \quad & e^{-j\theta}e^{j2\pi(f_i-f_c)\tau} \quad \cdot \quad e^{j2\pi\frac{\tau}{T_s}}H_0\left(f_i+\tfrac{1}{T_s}\right)x_0\left(f_i+\tfrac{1}{T_s}\right) \\
+ \quad & e^{j\theta}e^{j2\pi(f_i-f_c)\tau} \quad \cdot \quad H_0^*\left(2f_c-f_i\right)x_0^*\left(2f_c-f_i\right) \\
+ \quad & e^{j\theta}e^{j2\pi(f_i-f_c)\tau} \quad \cdot \quad e^{j2\pi\frac{\tau}{T_s}}H_0^*\left(2f_c-f_i-\tfrac{1}{T_s}\right)x_0^*\left(2f_c-f_i-\tfrac{1}{T_s}\right) \\
+ \quad & e^{-j\theta}e^{j2\pi(f_i-f_c)\tau} \quad \cdot \quad H_1\left(f_i\right)x_1\left(f_i\right) \\
+ \quad & e^{-j\theta}e^{j2\pi(f_i-f_c)\tau} \quad \cdot \quad e^{j2\pi\frac{\tau}{T_s}}H_1\left(f_i+\tfrac{1}{T_s}\right)x_1\left(f_i+\tfrac{1}{T_s}\right) \\
+ \quad & e^{j\theta}e^{j2\pi(f_i-f_c)\tau} \quad \cdot \quad H_1^*\left(2f_c-f_i\right)x_1^*\left(2f_c-f_i\right) \\
+ \quad & e^{j\theta}e^{j2\pi(f_i-f_c)\tau} \quad \cdot \quad e^{j2\pi\frac{\tau}{T_s}}H_1^*\left(2f_c-f_i-\tfrac{1}{T_s}\right)x_1^*\left(2f_c-f_i-\tfrac{1}{T_s}\right),
\end{aligned}
$$

for $f \in \left(f_c - \frac{1}{T_s}, f_c\right)$.

All that remains is to specify the forms of these filters and their associated equalizer. Multiplying out Eqn. (112), the optimal filters are,

$$
H_0\left(f\right) = \frac{1}{2}H_{\mathrm{EQ}}\left(f\right)\frac{A_0 S_{n,1}\left(f\right) - A_1 e^{-j\theta_d}e^{j2\pi f\tau_{d1}}S_{n,01}\left(f\right)}{S_{n,0}\left(f\right)S_{n,1}\left(f\right) - \left|S_{n,01}\left(f\right)\right|^2}\Psi^*\left(f\right) \tag{114}
$$

and

$$
H_1\left(f\right) = \frac{1}{2}H_{\mathrm{EQ}}\left(f\right)\frac{-A_0 S_{n,01}^*\left(f\right) + A_1 e^{-j\theta_d}e^{j2\pi f\tau_{d1}}S_{n,0}\left(f\right)}{S_{n,0}\left(f\right)S_{n,1}\left(f\right) - \left|S_{n,01}\left(f\right)\right|^2}\Psi^*\left(f\right). \tag{115}
$$
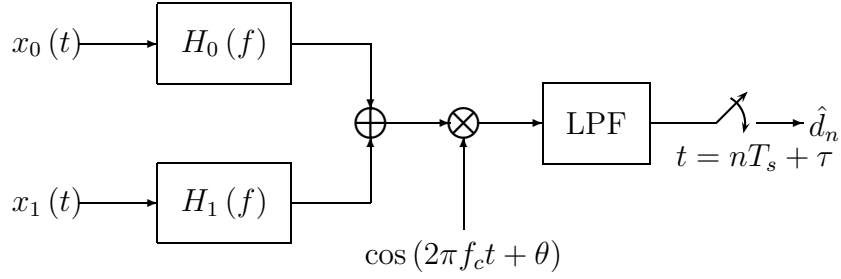
Figure 7.   Dual Sensor MMSE System Diagram

The equalizer, however, is much more complicated than the single sensor equalizer. In this case the equalizer, $H_{\text{EQ}}(f)$, is given by,

$$
\begin{aligned}
H_{\text{EQ}}(f) \;=\; & \frac{1}{\begin{aligned}1 \;+\; & \tfrac{1}{4}\left|\Psi(f-f_c)\right|^2 \boldsymbol{a}^\dagger \mathbf{R}_n^{-1}(f)\,\boldsymbol{a} \\ +\; & \tfrac{1}{4}\left|\Psi\left(f+\tfrac{1}{T_s}-f_c\right)\right|^2 \boldsymbol{a}^\dagger \mathbf{R}_n^{-1}\left(f+\tfrac{1}{T_s}\right)\boldsymbol{a} \\ +\; & \tfrac{1}{4}\left|\Psi(f_c-f)\right|^2 \boldsymbol{a}^\dagger \mathbf{R}_n^{-1}(2f_c-f)\,\boldsymbol{a} \\ +\; & \tfrac{1}{4}\left|\Psi\left(f_c-f-\tfrac{1}{T_s}\right)\right|^2 \boldsymbol{a}^\dagger \mathbf{R}_n^{-1}\left(2f_c-f-\tfrac{1}{T_s}\right)\boldsymbol{a}\end{aligned}}
\end{aligned}
\tag{116}
$$

where

$$
\boldsymbol{a}^\dagger \mathbf{R}_n^{-1}(f)\,\boldsymbol{a} \;=\; \frac{A_0^2 S_{n,1}(f) + A_1^2 S_{n,0}(f) - 2A_0 A_1 \Re\left\{e^{-j\theta_{d1}} e^{j2\pi f \tau_{di}} S_{n,01}(f)\right\}}{S_{n,0}(f)\,S_{n,1}(f) - \left|S_{n,01}(f)\right|^2}.
\tag{117}
$$

As before, this equation only specifies the filter for frequencies from $f_c - \frac{1}{T_s}$ to f_c. For frequencies from f_c to $f_c + \frac{1}{T_s}$, the sign of the $\frac{1}{T_s}$ terms needs to flip.

Implementing this filter in practice requires some small modifications to the system appropriate for a single sensor. In particular, each sensor is filtered separately and then the results are summed together. This sum is then down–converted and sampled as shown in Fig. 7.

Calculating the MSE at the output of this filter is similar to calculating the MSE at the output of the single sensor filter given in Eqn. (81). In this case the

MSE is,

$$\xi^2 \;\; = \;\; 2T_s \int_{f_c - \frac{1}{2T_s}}^{f_c} H_{\text{EQ}}\left(f\right) df. \tag{118}$$

Likewise the MCRB on the vector $\boldsymbol{d}$, according to this model,[7] is

$$MCRB\left(\boldsymbol{d}\right) \;\; = \;\; \left(\frac{1}{4}\boldsymbol{\Psi}_m^{\dagger}\mathbf{R}_{\Phi}^{\dagger}\mathbf{R}_n^{-1}\mathbf{R}_{\Phi}\boldsymbol{\Psi}_m + \mathbf{R}_d^{-1}\right)^{-1}. \tag{119}$$

This implies that further improvements in symbol estimation cannot be had without a better way of representing the true discrete probability distribution of a signal.

Further development, not presented here, could easily take this one step farther in order to specify what the MSE would be under arbitrary receiver conditions. Such a development would be nearly identical to that in Sec. 3.2.2, save that minor modifications would need to be made for the non–diagonal covariance matrix, $\mathbf{R}_n$.

In conclusion, this subsection has shown how the signal model first presented in Sec. 3.1 can be extended to derive multi–sensor filters for BPSK systems. Filters created using this model will be tested in Sec. 4.2.3 under extreme interference conditions. These simulations will demonstrate that the addition of even one additional sensor improves performance in highly correlated noise environments over the single sensor BPSK MMSE filter. Further, this gain holds even if the signal is only present on one sensor!

3.3   Estimating Time Difference of Arrival

Having developed several estimates of $\boldsymbol{d}$, we now proceed to the second application area of estimating the unknown Time Difference of Arrival (TDOA) between

---

[7]Modified Cramér–Rao bounds for the parameters τ_{di}, and A_i will be presented in the TDOA section, 3.3. As with the single sensor case, estimating these parameters does not reduce the theoretical capability of the receiver in terms of MSE.

two received signals. To estimate this TDOA, we again turn to "classical statistical principles," and in particular to the "maximum likelihood" estimation problem that others have found so difficult to accomplish in the past [23, p. 1177], Streight's recent work excepted [61]. Using these principles, this section derives the maximum likelihood function for estimating TDOA and then presents several practical approximations to that algorithm. Finally, this section concludes by examining the Cramér–Rao bound for TDOA estimation.

The maximum likelihood TDOA estimator is derived from a modified version of the model presented in Sec. 3.1 and in particular from Eqn. (52). In this modified model, we assume the signal subspace method applies for some initial, reference, sensor as before. All of the other sensors then have delayed versions of this signal relative to the reference. Thus the signal received on the reference sensor, sensor 0, can be written as

$$\boldsymbol{x}_{0,\text{SIGNAL}} \;=\; \frac{A_0}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}, \tag{120}$$

where the definitions of $\mathbf{R}_\phi$, $\boldsymbol{\Psi}$, and $\boldsymbol{d}$ remain unchanged from the initial signal model. The signal, as it arrives on the other sensors, differs from the reference in two respects. The first difference is that the other sensors may experience a different delay, τ_{di}, with respect to the reference. The second difference is a potential phase difference, defined as $\theta_{di} \triangleq \theta_i - \theta_0$, where θ_i is the received phase on the ith sensor. This dependence is captured by the matrix, $\mathbf{D}_i$, introduced in Eqn. (107) on page 76, which contains both delay difference and phase terms and is defined so that the received signal on the ith sensor may be written as,

$$\boldsymbol{x}_i \;=\; \frac{A_i}{2}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} + \boldsymbol{n}_i, \tag{121}$$

where $\mathbf{D}_0 = \mathbf{I}$ and where $\boldsymbol{x}_i$ is Gaussian distributed with mean, $\frac{A_i}{2}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}$, and variance $\mathbf{R}_{ni}$,

$$\boldsymbol{x}_i \sim \mathcal{N}\left(\frac{A_i}{2}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}, \mathbf{R}_{ni}\right). \tag{122}$$

A similar comment needs to be made regarding the noise probability distribution on multiple sensors. The following discussion assumes that the noise contributions to each sensor, $\boldsymbol{n}_i$, are uncorrelated,

$$\mathcal{E}\left\{\boldsymbol{n}_i\boldsymbol{n}_j^\dagger\right\}_{i\neq j} = \mathbf{0}. \tag{123}$$

This simplifies the overall covariance matrix, from block diagonal to purely diagonal.

The log likelihood of receiving signals from M sensors, $\boldsymbol{x}_0$ through $\boldsymbol{x}_{M-1}$, then separates into a sum,

$$\begin{aligned}
\mathcal{L} = \ & -\ \frac{mN_fM + N_f}{2}\ln(2\pi) - \frac{1}{2}\ln\det|\mathbf{R}_{\boldsymbol{d}}| - \frac{1}{2}\sum_{i=0}^{M-1}\ln\det|\mathbf{R}_{ni}| \tag{124}\\
& -\ \frac{1}{2}\sum_{i=0}^{M-1}\boldsymbol{x}_i^\dagger\mathbf{R}_{ni}^{-1}\boldsymbol{x}_i + \frac{A_i}{2}\Re\left\{\boldsymbol{x}_i^\dagger\mathbf{R}_{ni}^{-1}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right\} - \frac{A_i^2}{8}\boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{D}_i^\dagger\mathbf{R}_{ni}^{-1}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} \\
& -\ \frac{1}{2}\boldsymbol{d}^\dagger\mathbf{R}_d^{-1}\boldsymbol{d}.
\end{aligned}$$

Using the same method as before, the multi–sensor maximum likelihood estimate of $\boldsymbol{d}$ can be derived:

$$\hat{\boldsymbol{d}}_{\mathrm{MML}} = \left(\sum_{i=0}^{M-1}\frac{A_i^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_{ni}^{-1}\boldsymbol{\Psi} + \mathbf{R}_{\boldsymbol{d}}^{-1}\right)^{-1}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\sum_{k=0}^{M-1}\frac{A_k}{2}\mathbf{D}_k^\dagger\mathbf{R}_{nk}^{-1}\boldsymbol{x}_k. \tag{125}$$

If you plug this value into the log likelihood function, and drop all of the constant terms, the result is a new equation to maximize:

$$\hat{\boldsymbol{\tau}}_{d,\text{ML}} = \arg\max_{\boldsymbol{\tau}_d} \left(\sum_{i=0}^{M-1} \frac{A_i}{2} \boldsymbol{x}_i^\dagger \mathbf{R}_{ni}^{-1} \mathbf{D}_i \right) \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{H}_{\text{EQ}} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \left(\sum_{k=0}^{M-1} \frac{A_k}{2} \mathbf{D}_k^\dagger \mathbf{R}_{nk}^{-1} \boldsymbol{x}_k \right), \quad (126)$$

where

$$\mathbf{H}_{\text{EQ}} \triangleq \left(\sum_{k=0}^{M-1} \frac{A_k^2}{4} \boldsymbol{\Psi}^\dagger \mathbf{R}_{nk}^{-1} \boldsymbol{\Psi} + \mathbf{R}_{\boldsymbol{d}}^{-1} \right)^{-1}.$$

It is interesting to note that, in the case where no redundancy is present such that $\boldsymbol{\Psi}$ reduces to an $N_f \times N_f$ diagonal matrix and $\mathbf{R}_\phi$ is the identity, this equation reduces to the optimal filter for stationary TDOA estimation presented in Eqn. (24) of Sec. 2.3.1. This alone suggests that this estimator may have some optimality properties associated with it.

Unfortunately, this equation depends upon the unknown received signal gains inside a matrix inverse expression. Without simplifying this inverse in some manner, solving this system will require some amount of numerical iteration to determine the unknown gains, A_i. An alternate method, appropriate when the signal is weak on all sensors, would be to approximate $\mathbf{H}_{\text{EQ}}$ with $\mathbf{R}_d$. If this approximation is accepted, the low–SNR method for estimating the time–difference of arrival of a signal across M sensors is,

$$\hat{\boldsymbol{\tau}}_{d,\text{LOW-SNR}} = \arg\max_{\boldsymbol{\tau}_d} \left(\sum_{i=0}^{M-1} A_i \boldsymbol{x}_i^\dagger \mathbf{R}_{ni}^{-1} \mathbf{D}_i \right) \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \left(\sum_{k=0}^{M-1} A_k \mathbf{D}_k^\dagger \mathbf{R}_{nk}^{-1} \boldsymbol{x}_k \right).$$

$$(127)$$

At this point, rather than tackling this multi–variate optimization problem, it is worth noting that this estimate alone is a new result. While it is similar to Streight's low–SNR approximation to the maximum likelihood two sensor TDOA estimate given in Eqn. (29) [61], four differences separate the two estimators. The

first difference is that Streight's estimator was developed for white noise environments, where the covariance on the kth sensor, $\mathbf{R}_{nk}$, was proportional to the identity matrix, and then applied in colored noise environments, $\mathbf{R}_{nk} \not\propto \mathbf{I}$. It is reasonable to assume that applying the overwhitener portion of this estimator, $\mathbf{R}_{nk}^{-1}$, under colored noise conditions would provide a significant increase in performance. Second, while Streight's estimator expands into a sum of terms as well, $\sum_{i,j} \boldsymbol{x}_i^\dagger \cdots \boldsymbol{x}_j$, his estimator does not include the term containing $\boldsymbol{x}_0$ twice. While this term does not contain the "parameter of interest" [61, p. 75], τ_{di}, it does contain the nuisance parameters τ and θ that need to be estimated. Third, Streight assumes that the spectral correlation functions are unknown, whereas this estimator specifies them explicitly. The fourth difference is that this estimate can be compared with the optimal TDOA estimator under all SNR conditions given in Eqn. (126). Given that there are so many differences between the most similar two sensor cyclic TDOA estimator and this one, it seems prudent to first outline and validate these differences on the simpler, two–sensor problem and to reserve the multi–variate optimization problem for future work.

The rest of this chapter, then, is divided into three subsections. The first subsection is devoted to a thorough solution of the two sensor TDOA estimation problem. The second subsection, following the single cycle reasoning presented in [23], presents a single cycle approximation to this detector. Finally, the last subsection calculates several Modified Cramér–Rao Bounds (MCRB) appropriate for cyclic TDOA estimation.

3.3.1 Dual Sensor TDOA Estimation. This subsection describes a thorough solution of the two sensor TDOA estimation problem, focusing in particular on three parts that have not been addressed before. The first part is estimating the nuisance parameters, θ_{d1} and τ. Because the assumption has always been made that the spectral correlation function was entirely known, these angles have always been estimated as part of estimating the whole spectral correlation function. The

assumption here, that this correlation function is known, necessitates estimating θ_{d1} and τ for the first time. The second part of the two sensor problem is how to achieve subsample TDOA resolution. This results in an interpolation problem that has not been addressed before among cyclostationary TDOA estimators. Not interpolating, however, has resulted in misleading results in both [24] and [61] as Sec. 4.3.1 will demonstrate. The third part of the problem revolves around what to do with the unknown gain terms. All of these three parts to the TDOA estimation problem can be demonstrated with a QPSK signal alone, therefore this section will focus only on the redundancies present in a QPSK signal.[8]

Determining the nuisance parameters can be done by simplifying Eqn. (127) above. The first step is to eliminate the terms depending upon more than two sensors,

$$
\begin{aligned}
(\hat{\tau}_d)_{\text{2D,LOW-SNR}} \;=\; \arg\max_{\tau_d}\max_{\theta_d,\tau} \quad & A_0^2 \boldsymbol{x}_0^\dagger \mathbf{R}_{n,0}^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_{n,0}^{-1} \boldsymbol{x}_0 && (128) \\
+\; & A_1^2 \boldsymbol{x}_1^\dagger \mathbf{R}_{n,1}^{-1} \mathbf{D}_1 \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{D}_1^\dagger \mathbf{R}_{n,1}^{-1} \boldsymbol{x}_1 \\
+\; & 2 A_0 A_1 \Re \left\{ \boldsymbol{x}_0^\dagger \mathbf{R}_{n,0}^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{D}_1^\dagger \mathbf{R}_{n,1}^{-1} \boldsymbol{x}_1 \right\}.
\end{aligned}
$$

Then, plugging in the appropriate equations for a QPSK signal, and dropping the power terms that contain no information regarding either the TDOA parameter of

---

[8]Application to BPSK signals will require estimating the additional reference angle θ. It also includes 6 more cyclic spectral terms. Thus, considering QPSK only is a matter of simplification as well.

interest or the nuisance parameters, the maximization problem takes the form,

$$
\begin{aligned}
\hat{\tau}_d &= \arg\max_{\tau_d}\max_{\tau,\theta_d} A_0 A_1 \sum_{f_i} |\Psi(f_i - f_c)|^2 \, \Re\left\{ e^{-j\theta_d} \frac{X_0^*(f_i) X_1(f_i) e^{j2\pi f_i \tau_d}}{S_{n_0}(f_i) S_{n_1}(f_i)} \right\} \\
&+ A_0 A_1 \sum_{f_i} \left|\Psi\left(f_i + \frac{1}{T_s} - f_c\right)\right|^2 \Re\left\{ e^{-j\theta_d} \frac{X_0^*\left(f_i + \frac{1}{T_s}\right) X_1\left(f_i + \frac{1}{T_s}\right) e^{j2\pi\left(f_i + \frac{1}{T_s}\right)\tau_d}}{S_{n_0}\left(f_i + \frac{1}{T_s}\right) S_{n_1}\left(f_i + \frac{1}{T_s}\right)} \right\} \\
&+ \Re\left\{ e^{j2\pi\frac{\tau}{T_s}} \sum_{f_i} \Psi(f_i - f_c)\Psi^*\left(f_i + \frac{1}{T_s} - f_c\right) \left[\begin{array}{l} A_0^2 \frac{X_0^*(f_i) X_0\left(f_i + \frac{1}{T_s}\right)}{S_{n_0}(f_i) S_{n_0}\left(f_i + \frac{1}{T_s}\right)} \\ + A_1^2 \frac{X_1^*(f_i) X_1\left(f_i + \frac{1}{T_s}\right)}{S_{n_1}(f_i) S_{n_1}\left(f_i + \frac{1}{T_s}\right)} e^{j2\pi\frac{\tau_d}{T_s}} \\ + A_0 A_1 e^{-j\theta_d} \frac{X_0^*(f_i) X_1\left(f_i + \frac{1}{T_s}\right)}{S_{n_0}(f_i) S_{n_1}\left(f_i + \frac{1}{T_s}\right)} e^{j2\pi\left(f_i + \frac{1}{T_s}\right)\tau_d} \\ + A_0 A_1 e^{j\theta_d} \frac{X_1^*(f_i) X_0\left(f_i + \frac{1}{T_s}\right)}{S_{n_1}(f_i) S_{n_0}\left(f_i + \frac{1}{T_s}\right)} e^{-j2\pi f_i \tau_d} \end{array} \right] \right\}.
\end{aligned}
$$

$$(129)$$

Several unknowns persist in this equation. These are the TDOA parameter, τ_d, together with the nuisance parameters A_0, A_1, θ_d, and τ. In order to separate the problem of solving for the TDOA, τ_d, from the related problem of solving for the other parameters, we break this equation into parts corresponding to functions of τ_d and functions of the nuisance parameters. This yields the expression,

$$
\hat{\tau}_d = \arg\max_{\tau_d}\max_{\tau,\theta_d} \Re\left\{ \begin{array}{l} A_0 A_1 e^{-j\theta_d} B_0[\tau_d] \\ + A_0 A_1 e^{-j\theta_d} e^{j2\pi\frac{\tau}{T_s}} B_1[\tau_d] \\ + A_0 A_1 e^{-j\theta_d} e^{-j2\pi\frac{\tau}{T_s}} B_2[\tau_d] \\ + A_0^2 e^{j2\pi\frac{\tau}{T_s}} B_3 \\ + A_1^2 e^{j2\pi\frac{\tau}{T_s}} e^{j2\pi\frac{\tau_d}{T_s}} B_4 \end{array} \right\}, \tag{130}
$$

86

where the B_i terms are defined as,

$$B_0\left[\tau_d\right] \triangleq \sum_{f_i}\left[\begin{array}{c} \frac{|\Psi(f_i-f_c)|^2}{S_{n,0}(f_i)S_{n,1}(f_i)}X_1\left(f_i\right)X_0^*\left(f_i\right)e^{j2\pi f_i\tau_d} \\ + \frac{\left|\Psi\left(f_i+\frac{1}{T_s}-f_c\right)\right|^2}{S_{n,0}\left(f_i+\frac{1}{T_s}\right)S_{n,1}\left(f_i+\frac{1}{T_s}\right)}X_1\left(f_i+\frac{1}{T_s}\right)X_0^*\left(f_i+\frac{1}{T_s}\right)e^{j2\pi\left(f_i+\frac{1}{T_s}\right)\tau_d} \end{array}\right], \tag{131}$$

$$B_1\left[\tau_d\right] \triangleq \sum_{f_i}\frac{\Psi\left(f_i-f_c\right)\Psi^*\left(f_i+\frac{1}{T_s}-f_c\right)}{S_{n,0}\left(f_i\right)S_{n,1}\left(f_i+\frac{1}{T_s}\right)}X_0^*\left(f_i\right)X_1\left(f_i+\frac{1}{T_s}\right)e^{j2\pi\left(f_i+\frac{1}{T_s}\right)\tau_d}, \tag{132}$$

$$B_2\left[\tau_d\right] \triangleq \sum_{f_i}\frac{\Psi^*\left(f_i-f_c\right)\Psi\left(f_i+\frac{1}{T_s}-f_c\right)}{S_{n,0}\left(f_i+\frac{1}{T_s}\right)S_{n,1}\left(f_i\right)}X_0^*\left(f_i+\frac{1}{T_s}\right)X_1\left(f_i\right)e^{j2\pi f_i\tau_d}, \tag{133}$$

$$B_3 \triangleq \sum_{f_i}\frac{\Psi\left(f_i-f_c\right)\Psi^*\left(f_i+\frac{1}{T_s}-f_c\right)}{S_{n,0}\left(f_i\right)S_{n,0}\left(f_i+\frac{1}{T_s}\right)}X_0^*\left(f_i\right)X_0\left(f_i+\frac{1}{T_s}\right), \tag{134}$$

$$\text{and } B_4 \triangleq \sum_{f_i}\frac{\Psi\left(f_i-f_c\right)\Psi^*\left(f_i+\frac{1}{T_s}-f_c\right)}{S_{n,1}\left(f_i\right)S_{n,1}\left(f_i+\frac{1}{T_s}\right)}X_1^*\left(f_i\right)X_1\left(f_i+\frac{1}{T_s}\right). \tag{135}$$

Of these new B_i parameters, the B_0 parameter should be familiar: it corresponds to TDOA estimation using the Eckart filter [37]. The other parameters are not as familiar, as cyclostationary developments have generally assumed that $\Psi\left(f\right)$ and $S_n\left(f\right)$ were unknown. Maximizing Eqn. (130) will provide the maximum likelihood TDOA estimate. Even better, the B_i expressions can be calculated without knowing the nuisance parameters.

In order to apply this formula in practice, the first step to calculating the maximum of Eqn. (130) is to calculate the B_i functions. The first three of these, $B_0, \ldots, B_2$ are functions of τ_d and are calculated by three inverse Fast Fourier Transforms (FFT) [48]. The last two, B_3 and B_4, can be calculated by simply applying the sum in Eqns. (134) and (135).

Once these functions have been calculated, maximizing Eqn. (130) becomes a problem of solving for the nuisance parameters at each value of τ_d. The first nuisance

parameter to estimate is θ_d. Noting that, in general, $\left|B_0\left[\tau_d\right]\right|$ will be much larger than $\left|e^{j2\pi\frac{\tau}{T_s}}B_1\left[\tau_d\right]+e^{-j2\pi\frac{\tau}{T_s}}B_2\left[\tau_d\right]\right|$, a simple estimate for $e^{j\theta_d}$ can be calculated from this parameter alone,

$$\zeta\left[\tau_d\right] \triangleq e^{-j\hat{\theta}_d} = \frac{B_0^*\left[\tau_d\right]}{\left|B_0\left[\tau_d\right]\right|}. \tag{136}$$

Using this estimate for $e^{j\hat{\theta}_d}$ to simplify the optimization problem further,

$$\hat{\tau}_d = \arg\max_{\tau_d}\left|B_0\left[\tau_d\right]\right| \tag{137}$$
$$+\max_{\tau}\Re\left\{e^{j2\pi\frac{\tau}{T_s}}\left[\zeta\left[\tau_d\right]B_1\left[\tau_d\right]+\zeta^*\left[\tau_d\right]B_2^*\left[\tau_d\right]+\frac{A_0}{A_1}B_3+\frac{A_1}{A_0}e^{j2\pi\frac{\tau_d}{T_s}}B_4\right]\right\}.$$

At this point, since the $e^{j2\pi\frac{\tau}{T_s}}$ term can be factored out, the optimal estimate is apparent,

$$\hat{\tau}_d = \arg\max_{\tau_d}\left|B_0\left[\tau_d\right]\right|+\left|\zeta\left[\tau_d\right]B_1\left[\tau_d\right]+\zeta^*\left[\tau_d\right]B_2^*\left[\tau_d\right]+\frac{A_0}{A_1}B_3+\frac{A_1}{A_0}e^{j2\pi\frac{\tau_d}{T_s}}B_4\right|. \tag{138}$$

This solves for all of the unknown nuisance angles, yielding a very usable TDOA estimator.

This opens up the second part of this two–sensor problem: what happens when τ_d is not an integer number of samples? The most obvious answer is to apply some form of interpolation to the function in Eqn. (138) and to use that to determine fractional delays. To do this, we define the function $g\left[\tau_d\right]$, to be the right hand side of Eqn. (138) sampled at intervals of T_{samp},

$$g\left[\tau_d\right] \triangleq \left|B_0\left[\tau_d\right]\right|+\left|\zeta\left[\tau_d\right]B_1\left[\tau_d\right]+\zeta^*\left[\tau_d\right]B_2^*\left[\tau_d\right]+\frac{A_0}{A_1}B_3+\frac{A_1}{A_0}e^{j2\pi\frac{\tau_d}{T_s}}B_4\right|. \tag{139}$$

Maximizing $g[\tau_d]$ involves two steps. The first step is a search for the maximum over all sampled values of $g[\tau_d]$. Once found, some form of local interpolation near that maximum can be used to achieve subsample resolution. The form of the optimal interpolator is easily given from the forms of $B_0, \ldots, B_2$. Each of these functions is defined as a continuous function of τ_d, not a discrete one. By increasing the size of the inverse FFT used to calculate these functions, subsample resolution may be obtained. However, this method is computationally intensive and could be pursued *ad infinitum*. A second, non–optimal, option would be to use some form of polynomial interpolation to achieve sub–sample resolution. In that case, the value of τ_d that produced a maximum in $g[\tau_{d1}]$, call this $\hat{\hat{\tau}}_d$, would then be applied to the quadratic interpolation formula given in [5] to determine an optimal TDOA value,

$$\hat{\tau}_d = \hat{\hat{\tau}}_d + \frac{1}{2} \frac{g\left[\hat{\hat{\tau}}_d + T_{\text{samp}}\right] - g\left[\hat{\hat{\tau}}_d - T_{\text{samp}}\right]}{2g\left[\hat{\hat{\tau}}_d\right] - \left(g\left[\hat{\hat{\tau}}_d - T_{\text{samp}}\right] + g\left[\hat{\hat{\tau}}_d + T_{\text{samp}}\right]\right)}. \tag{140}$$

Perhaps the best option, however, is a combination of a larger FFT size together with quadratic interpolation. This combination method will be tested in Chapt. IV.

The final problem of TDOA MLE implementation is what to do about the $\frac{A_0}{A_1}$ term in Eqn. (139). This problem can be avoided by one assumption and two approximations. First, assume that it is known that $A_0 \gg A_1$. In this case, $\frac{A_1}{A_0} B_4 e^{j2\pi \frac{\tau_d}{T_s}}$ is a small term that may be dropped with little loss from Eqn. (139). The remaining term dependent on $\frac{A_0}{A_1}$ is dealt with by applying a less than optimal scale. This creates a lopsided approximation to the TDOA optimization function that is appropriate when the signal scales are unequal,

$$\tilde{g}[\tau_d] \triangleq \left|B_0[\tau_d]\right| + \left|\zeta[\tau_d] B_1[\tau_d] + \zeta^*[\tau_d] B_2^*[\tau_d] + B_3\right|. \tag{141}$$

In all other respects, such as interpolation, this lopsided function is optimized the same as before.

In sum, this section has presented three new techniques to be used in TDOA estimation. The first new technique, presented in Eqn. (138), demonstrates a method of dealing with the unknown nuisance parameters. Since this is the first work to assume, in TDOA estimation, that these spectral correlation functions are known save these nuisance parameters, this nuisance parameter estimation problem makes a cyclic TDOA estimator practical. Further, it renders the problem of solving for the spectral correlation functions unnecessary, thus simplifying the problem. The second technique presented here, interpolating the likelihood function between sample points, is necessary whenever the true TDOA is not necessarily an integer number of samples. Since this includes all real world cases, applying this interpolation should improve all practical results. Finally, a method for estimating TDOA when signal scales were unknown was presented in Eqn. (141). Together, these methods cover all of the problems associated with unknowns in the spectral correlation functions making maximum likelihood cyclic TDOA estimation practical.

3.3.2 Single Cycle TDOA Estimators.   Under the justification that co–channel interference would corrupt the $B_0[\tau_d]$ from Eqn. (131), Gardner and Chen focus on several *single cycle* TDOA estimators [23]. These estimators, they argue, are immune to unknown interference after sufficient integration lengths because,

> By exploiting the cyclostationary property of the signal of interest, as reflected in the spectral correlation functions for the received data, the effects of additive noise and interfering signals are ideally (for unlimited data collection times) removed by these methods. [23, p. 1182]

Since all of these methods focus around a single $\alpha \neq 0$ cycle frequency, this subsection examines single–cycle TDOA estimators in light of the maximum likelihood TDOA estimation function. Two particular single cycle estimators are presented here. The first is a modified spectral coherence alignment method, M–SPECCOA, modified

here to accommodate known colored noise environments. The second single cycle estimator is developed in this section by dropping the $B_0[\tau_d]$ term from the ML TDOA estimator. Both of these estimators, according to the philosophy in the quote above, should be highly resistant to noise plus interference. Since the signal of interest throughout this TDOA section has been a QPSK signal, only the symbol rate cycle frequency, $\alpha = \frac{1}{T_s}$, will be examined.

The first estimator of interest is SPECCOA. Under Gardner and Chen's formulation, this estimator is created from two estimated spectral correlation functions, $\hat{S}_{x_0 x_1}^{\alpha}(f)$ and $\hat{S}_{x_1 x_1}^{\alpha}(f)$. If the pulse function is known, however, the second spectral correlation function is known to within a complex constant. That is, for $\alpha = \frac{1}{T_s}$,

$$
S_{x_1 x_1}^{\frac{1}{T_s}}(f) = e^{-j2\pi \frac{\tau + \tau_d}{T_s}} \frac{A_0^2}{4T_s} \Psi^* \left(f - \frac{1}{2T_s} - f_c \right) \Psi \left(f + \frac{1}{2T_s} - f_c \right). \tag{142}
$$

Replacing the unknown, but estimated, value of $S_{x_1 x_1}^{\frac{1}{T_s}}(f)$ with its true value results in the single cycle estimator,

$$
\hat{\tau}_{d1,\text{NO-APPRX}} = \arg\max_{\tau_d} \left| e^{-j2\pi \frac{\tau + \tau_d}{T_s}} \int_{f_c - \frac{1}{2T_s}}^{f_c + \frac{1}{2T_s}} \begin{array}{c} \frac{A^2}{4T_s} \Psi^* \left(f - \frac{1}{2T_s} - f_c \right) \Psi \left(f + \frac{1}{2T_s} - f_c \right) \\ \times \quad X_1^* \left(f - \frac{1}{2T_s} \right) X_0 \left(f + \frac{1}{2T_s} \right) df \end{array} \right|
$$

$$
= \arg\max_{\tau_d} \left| B_1[\tau_d] \right|, \tag{143}
$$

in white noise conditions. Yet SPECCOA uses more signal information than this equation captures. In particular, the $S_{x_0 x_0}^{\frac{1}{T_s}}(f)$ term formed from $X_0^* \left(f - \frac{1}{2T_s} \right) X_0 \left(f + \frac{1}{2T_s} \right)$, contains information about $e^{-j2\pi \frac{\tau + \tau_d}{T_s}}$ not captured in this formula. Putting these two terms together results in a modified SPECCOA

estimator,

$$\hat{\tau}_{d,\text{M--SPECCOA}} \quad \triangleq \quad \arg\max_{\tau_d} \Re\left\{ e^{j\theta_d} B_1\left[\tau_d\right] B_3^* \right\}, \tag{144}$$

that captures most, if not all, of the information that the previous SPECCOA estimator captures. If the phase is unknown, the real operator, $\Re\{\cdot\}$ operator may be replaced with the absolute value operator, $|\cdot|$. Doing so, however, eliminates the contribution of B_3,

$$\hat{\tau}_{d,\text{M}^2\text{--SPECCOA}} \quad \triangleq \quad \arg\max_{\tau_d} \left| e^{j\theta_d} B_1\left[\tau_d\right] \right|, \tag{145}$$

since $|B_3|$ is constant across all values of τ_d.

Two differences separate Eqn. (144) from the original definition of SPECCOA. The first difference is that all of the terms in Eqn. (144) are well defined. Estimating $S_s^\alpha(f)$ is not required, as Eqn. (144) assumes that it is known. The second difference is found in colored noise environments. In such environments, B_1 and B_3 specify using an overwhitener prior to estimating τ_d. The importance of this overwhitener will be demonstrated in Sec. 4.3.2, where overwhitened and non–overwhitened versions of this estimator will be compared.

Realizing that this expression is just a subset of the maximization problem presented in the last section in Eqn. (130), it becomes apparent that much more single cycle information is present in the full maximum likelihood problem than SPECCOA uses. Applying all of this information should yield better single cycle TDOA estimators.

In an effort to use all of the terms $B_1, \ldots, B_4$, two more single cycle TDOA estimators are presented here. The first is an optimal single cycle method for use

when θ_d is unknown,

$$\hat{\tau}_{d,\text{SC-THETA}} \quad \triangleq \quad \arg\max_{\tau_d} \left\{ \left| B_1 \left[\tau_d \right] \right| + \left| B_2 \left[\tau_d \right] \right| \right\}. \tag{146}$$

The problem associated with not knowing θ_d means that the angles of these two terms cannot be related to each other. Therefore, the absolute value sign allows the TDOA estimator to maximize against both θ_d and τ in solving for τ_d. The second single cycle estimator, shown below, is optimal when $\theta_d = 0$. This estimator uses only $B_1, \ldots, B_3$ for the same reasons as the lopsided TDOA estimator above. In this case,

$$\hat{\tau}_{d,\text{SINGLE-CYCLE}} \quad \triangleq \quad \arg\max_{\tau_d} \left| B_1 \left[\tau_d \right] + B_2^* \left[\tau_d \right] + B_3 \right| \tag{147}$$

is a suboptimal single cycle TDOA estimator for use when A_0 and A_1 are unknown, but $A_0 \gg A_1$.

Two of these single cycle TDOA estimators, the modified SPECCOA in Eqn. (144) and the optimal single cycle method when $A_0 \gg A_1$ in Eqn. (147), will be tested in Section 4.3 of the next chapter. This section will demonstrate that their performance is much worse than the TDOA estimators presented in the previous subsection which incorporate the $B_0 \left[\tau_d \right]$, or zero–cycle, term—even in severe co–channel interference. This poor performance, demonstrated in Chapt. IV, leaves little reason to discuss them further here.

3.3.3 Cramér–Rao Bounds. As a last step in discussing cyclic TDOA estimation, we return to the modified Cramér–Rao bound discussed in Sec. 3.2.3. Deriving this bound for the multi–sensor likelihood function that we've been using is fairly straightforward but lengthy. A full presentation of this derivation can be found in Appendix J. Instead, this section discusses the derivation in order to present

results for three particular cases. The first two cases assume that τ, the reference symbol epoch, is the only unknown nuisance parameter. These two cases present an MCRB for τ_{d1} when $f_c = 0$ and then when $f_c \gg 0$. The third case occurs when both nuisance parameters, τ and θ_d are unknown and $f_c \gg 0$. In all of these cases, as with the previous TDOA algorithm development, a QPSK signal will be the basis for the model.

The first step in any MCRB calculation involves calculating the expected values of the second partial of the likelihood function. For the two sensor case, this second partial involves taking partials of $\mathbf{D}_i$ with respect to τ_{d1}. To simplify this process, a new matrix, $\mathbf{F}_\Delta$, is introduced here. This matrix is similar to $\mathbf{F}_\delta$ (Eqn. (98)), introduced in Sec. 3.2.3 on page 73, only here it is applied to $\mathbf{D}_i$ instead of $\mathbf{R}_\phi$. Its definition shows that similarity,

$$\mathbf{F}_\Delta \triangleq \frac{1}{-j2\pi}\left(\frac{\partial}{\partial\tau_{di}}\mathbf{D}_i\right)\mathbf{D}_i^\dagger = \begin{bmatrix} f_1 & 0 & \cdots \\ 0 & f_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{148}$$

Moreover, when $f_c = 0$, these two auxiliary matrices are identical, $\mathbf{F}_\delta = \mathbf{F}_\Delta$.

In order to deal with the first case, where τ is unknown and $f_c = 0$, one minor modification needs to be made to Eqn. (1). A baseband QPSK signal is complex, not real. To put it another way,

$$x_{\text{BB}}(t) = A_0 \sum_{n=0}^{N_s} d_n \psi(t - nT_s - \tau) e^{j\theta} + n(t). \tag{149}$$

The biggest immediate consequence of this change is that there is no longer a factor of two scaling in the vector equation, $\boldsymbol{x} = \frac{A}{2}\mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} + \boldsymbol{n}$. Instead, we have

$$\boldsymbol{x}_{\text{BB}} = A\mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} + \boldsymbol{n}, \tag{150}$$

$$\text{and } \boldsymbol{x}_{i,\text{BB}} = A_i \mathbf{D}_i \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} + \boldsymbol{n}_i. \tag{151}$$

Once this change is made, the Fisher Information Matrix (FIM) is fairly easy to derive. Three terms are of particular interest. These are,

$$\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau^2}\right\} = \sum_{k=0}^{M} 4\pi^2 A_k^2 \text{tr}\left\{\boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_{nk}^{-1} \mathbf{F}_\delta^2\right\}, \tag{152}$$

$$\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{di}^2}\right\} = 4\pi^2 A_i^2 \text{tr}\left\{\boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\delta^2\right\}, \tag{153}$$

$$\text{and } \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{di}\partial \tau}\right\} = 4\pi^2 A_i^2 \text{tr}\left\{\boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\delta^2\right\}. \tag{154}$$

From these three terms, the FIM is easily inverted for an arbitrary number of sensors. This leads to the following Modified Cramér–Rao Bounds (MCRBs),

$$MCRB\left(\tau\right) = \frac{1}{4\pi^2 A_0^2 \text{tr}\left\{\boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_{n0}^{-1} \mathbf{F}_\delta^2\right\}}, \tag{155}$$

$$\text{and } MCRB\left(\tau_{di}\right) = \frac{1}{4\pi^2 A_0^2 \text{tr}\left\{\boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_{n0}^{-1} \mathbf{F}_\delta^2\right\}} + \frac{1}{4\pi^2 A_i^2 \text{tr}\left\{\boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\delta^2\right\}}. \tag{156}$$

This bound is plotted in Sec. 4.3.1 together with the performance of several ML estimators. (See Figs. 37 through 40.)

The second case, where $f_c \gg 0$ and τ is the only unknown nuisance parameter, is not quite so easy. This is primarily due to the fact that $\mathbf{F}_\delta \neq \mathbf{F}_\Delta$, and so the multiple sensor MCRB is not as simply stated. Instead, this case will examine the two sensor CRB for QPSK signals rather than the multi–sensor case in general.

For two sensors, the relevant FIM terms are,

$$\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau^2}\right\} = \sum_{k=0}^{M} \pi^2 A_k^2 \text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger \mathbf{R}_{nk}^{-1}\mathbf{F}_\delta^2\right\}, \tag{157}$$

$$\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{di}^2}\right\} = \pi^2 A_i^2 \text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger \mathbf{R}_{ni}^{-1}\mathbf{F}_\Delta^2\right\}, \tag{158}$$

$$\text{and } \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{di}\partial \tau}\right\} = \pi^2 A_i^2 \text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger \mathbf{R}_{ni}^{-1}\mathbf{F}_\delta\mathbf{F}_\Delta\right\}. \tag{159}$$

When the FIM is inverted, these terms result in the bounds,

$$MCRB\left(\tau\right) = \frac{1}{\pi^2}\frac{1}{A_0^2\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{n0}^{-1}\mathbf{F}_\delta^2\right\} + A_1^2\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\mathbf{F}_\delta^2\right\} - \frac{A_1^2\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\mathbf{F}_\delta\mathbf{F}_\Delta\right\}^2}{\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\mathbf{F}_\Delta^2\right\}}}, \tag{160}$$

$$\text{and } MCRB\left(\tau_{d1}\right) = \frac{1}{\pi^2}\frac{1}{A_1^2\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\mathbf{F}_\Delta^2\right\} - \frac{A_1^4\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\mathbf{F}_\delta\mathbf{F}_\Delta\right\}^2}{A_0^2\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\mathbf{F}_\delta^2\right\}+A_1^2\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\mathbf{F}_\delta^2\right\}}}. \tag{161}$$

This MCRB for TDOA estimation is plotted in Sec. 4.3.2 (Figs. 43 and 45) for two colored noise test cases.

The final case, that where $f_c \gg 0$ and where τ and θ_{d1} must both be estimated, is a little more tedious to present since it involves inverting an arbitrary 3×3 matrix. Rather than presenting the bounds in this case, only the entries in the FIM will be shown here. From these entries, the bounds are easily calculated numerically. These entries are identical to the last case, with the exception that one more row needs to

be added to the matrix. This row is composed of the terms,

$$\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d1}^2}\right\} = \frac{A_1^2}{4}\mathrm{tr}\left\{\mathbf{\Psi}\mathbf{R}_d\mathbf{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\right\}, \tag{162}$$

$$\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d1} \partial \tau}\right\} = -\frac{A_1^2 \pi}{2}\mathrm{tr}\left\{\mathbf{\Psi}\mathbf{R}_d\mathbf{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\mathbf{F}_\delta\right\}, \tag{163}$$

$$\text{and } \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d1} \partial \tau_{d1}}\right\} = -\frac{A_1^2 \pi}{2}\mathrm{tr}\left\{\mathbf{\Psi}\mathbf{R}_d\mathbf{\Psi}^\dagger\mathbf{R}_{n1}^{-1}\mathbf{F}_\Delta\right\}. \tag{164}$$

Unlike the previous two cases, plots of this latter bound are not shown in the next chapter at all.

Further cases could be presented here for numerical solution. In particular, the MCRBs for an arbitrary number of sensors are easily calculated from a single matrix inverse. However, since these bounds are tedious to show analytically, they have not been included here. An interested reader can find the FIM entries in Appendix J and numerically invert them as desired.

3.4 Presence Detection

The same subspace framework that was used to create first optimal filters, and then TDOA estimators, can also be used to develop detection algorithms. Two hypotheses will be examined in this section to test for the presence of a signal. The first hypothesis test is the standard noise alone versus signal present test. This will lead to the well known multicycle detector. The second hypothesis tests whether or not the signal is present when the noise has an unknown scale. This test will result in a new *cyclic ratio* detector. Approximations to this latter detector will result in a signal selective detection capability that is much more resistant to interference than classical cyclostationary detection methods.

Making this problem more difficult than the optimal filtering problem is the fact that several of the values used to determine $\hat{\boldsymbol{d}}_{\mathrm{MLE}}$ are unknown. Chief among these unknowns is the signal scale parameter, A. Thus, rather than starting with

the likelihood ratio test, which is the most powerful test under known conditions, we derive only the *locally* most powerful test for low SNR conditions and approximations to that test. This test uses a detection statistic defined as,

$$y_{\mathrm{LO}}\left(\boldsymbol{x}\right) \quad = \quad \left.\frac{\partial}{\partial A^2}\mathcal{L}\right|_{A^2=0}. \tag{165}$$

This is the most powerful test for detecting a weak signal, such as one where $A^2 \approx 0$ [13].

3.4.1 Optimal Cyclostationary Signal Detection. The first test of interest is whether or not the signal is present when the noise covariance is known. An appropriate test can be derived from the hypotheses,

$$H_0 \quad : \quad \boldsymbol{X} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_n\right) \tag{166}$$

$$\text{versus } H_1 \quad : \quad \boldsymbol{X} \sim \mathcal{N}\left(\frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}, \mathbf{R}_n\right) \text{ and } \boldsymbol{d} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_d\right). \tag{167}$$

Calculating the locally most powerful low SNR detector for this signal, using Eqns. (165) and (70), yields the detection statistic,

$$y_{\mathrm{MCYC}}\left(\boldsymbol{x}\right) \quad = \quad \max_{\boldsymbol{\phi}}\boldsymbol{x}^\dagger\mathbf{R}_n^{-1}\mathbf{R}_\phi\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\boldsymbol{x}. \tag{168}$$

This detection statistic reduces to the multicycle detector presented by Gardner [17] and modified for colored noise by Rostaing [51] (see Eqn. (38) on page 38). This should come as no surprise, since this detector was derived under identical conditions to those under which the multicycle detector was derived [17, 15].

3.4.2 Cyclic Ratio Detection. The second test of interest revolves around determining whether or not the signal is present in burst interference. As mentioned in the background subsection on detection (sec. 2.4.1 on page 33), the biggest

drawback to energy detection methods is that they will indiscriminately detect all burst signals. Overcoming this difficulty requires deriving a signal detector that is somehow immune to burst interference. Using the signal model presented earlier in this chapter, together with an appropriately framed hypothesis test, it is possible to derive a detector that will be resistant to burst interference. This subsection presents the highlights of such a derivation. An interested reader may wish to refer to Appendix G for a more detailed proof.

The first step is to appropriately frame the test. We are interested in testing whether or not the signal is present in a background where the noise covariance could change suddenly. The difficult part in deriving such a test is that the noise covariance could change in any manner, yet in order to derive a test some manner of change needs to be specified. While one might be able to derive a test which is invariant to a particular type of burst interferer, such a test would need to be redesigned when the burst interferer changed. Instead, we choose here to be as general as possible by allowing the noise scale to change suddenly. Then, in Sec. 4.4, this result is shown to apply even when the shape of the noise PSD changes. Therefore, we let σ refer to this unknown and changing scale, and frame the hypothesis test as

$$H_0 \quad : \quad \boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2 \mathbf{R}_n\right) \tag{169}$$

$$\text{versus } H_1 \quad : \quad \boldsymbol{X} \sim \mathcal{N}\left(\frac{A}{2}\mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d}, \sigma^2 \mathbf{R}_n\right), \, A > 0, \text{ and } \boldsymbol{d} \sim \mathcal{N}\left(\boldsymbol{0}, \mathbf{R}_d\right). \tag{170}$$

From the previous section on filtering, Sec. 3.2.1, the maximum likelihood estimate for $\boldsymbol{d}$ is,

$$\hat{\boldsymbol{d}}_{\text{MLE}} \quad = \quad \frac{A}{2\sigma^2}\left(\frac{A^2}{4\sigma^2}\boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)^{-1}\boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1}\boldsymbol{x}. \tag{171}$$

Similarly, the maximum likelihood estimate for σ, given that $\boldsymbol{d}$ is known, is also easy to derive,

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{mN_f} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right)^\dagger \mathbf{R}_n^{-1} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right). \tag{172}$$

A proof of this may be found in [53]. The next step is to place these estimates into the likelihood function to remove them as unknowns. Yet the problem with these estimates is that they are mutually dependent. That is, the estimate for $\hat{\sigma}_{\text{MLE}}$ depends upon the estimate for $\hat{\boldsymbol{d}}_{\text{MLE}}$, and likewise $\hat{\boldsymbol{d}}_{\text{MLE}}$ depends upon $\hat{\sigma}_{\text{MLE}}$. They cannot be analytically separated.

To resolve this difficulty, we define a relative signal scale factor, A_o, such that $A_o \triangleq \frac{A}{2\sigma}$. This makes it possible to sufficiently decouple the estimates,

$$\hat{\boldsymbol{d}}_{\text{MLE}} = \frac{A_o}{\sigma} \left(A_o^2 \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}, \tag{173}$$

$$\text{and } \hat{\sigma}^2_{\text{MLE}} = \frac{1}{mN_f} \left[\begin{array}{l} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \\ - \quad 2A_o^2 \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(A_o^2 \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \\ + \quad A_o^4 \cdots \end{array} \right]. \tag{174}$$

While σ does remain in the expression for $\hat{\boldsymbol{d}}_{\text{MLE}}$, this dependence is dropped when $\hat{\boldsymbol{d}}_{\text{MLE}}$ is placed into the likelihood function.

Having estimated the unknown parameters in the likelihood function, these parameters can now be plugged in and the log of the likelihood function, $\mathcal{L}$, can be evaluated. If all the terms that are independent of the data are lumped into a constant, C, then the expression for $\mathcal{L}$ becomes,

$$\mathcal{L} = -\frac{mN_f}{2} \ln \hat{\sigma}^2_{\text{MLE}} - \frac{mN_f \hat{\sigma}^2_{\text{MLE}}}{2\hat{\sigma}^2_{\text{MLE}}} - \frac{\hat{\boldsymbol{d}}^\dagger_{\text{MLE}} \mathbf{R}_d^{-1} \hat{\boldsymbol{d}}_{\text{MLE}}}{2} + C. \tag{175}$$

Finally, using this likelihood, we develop a detector via the formula for locally optimal detection as shown in Eqn. (165) above. This detector is,

$$y_{\text{RATIO}}\left(\boldsymbol{x}\right) \quad \triangleq \quad \max_{\tau,\theta} \frac{\boldsymbol{x}^{\dagger}\mathbf{R}_n^{-1}\mathbf{R}_{\phi}\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^{\dagger}\mathbf{R}_{\phi}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{x}}{\boldsymbol{x}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{x}}. \tag{176}$$

Because of its form, the term *Cyclic Ratio Detector* is applied here to describe this detector. In particular, the term *multicycle ratio detector* seems appropriate since the numerator is still the optimal multicycle detector.

Before leaving this topic, two single cycle approximations are presented to this detector, creating *single cycle ratio detectors.* These detectors result when all but one of the terms is dropped from the numerator. Two important single cycle ratio detectors will be evaluated through simulation in Sec. 4.4. These are the symbol rate ratio detector,

$$y_{\text{SRD}}\left(\boldsymbol{x}\right) \quad \triangleq \quad \frac{\left|\displaystyle\sum_{f_i} \frac{X^*\left(f_i - \frac{1}{2T_s}\right)}{S_n\left(f_i - \frac{1}{2T_s}\right)} \frac{X\left(f_i + \frac{1}{2T_s}\right)}{S_n\left(f_i + \frac{1}{2T_s}\right)} S_s^{\frac{1}{T_s}}\left(f_i\right)^*\right|}{\displaystyle\sum_{f_i} \frac{|X\left(f_i\right)|^2}{S_n\left(f_i\right)} + \frac{\left|X\left(f_i + \frac{1}{T_s}\right)\right|^2}{S_n\left(f_i + \frac{1}{T_s}\right)}} \tag{177}$$

and the carrier ratio detector,

$$y_{\text{CRD}}\left(\boldsymbol{x}\right) \quad \triangleq \quad \frac{\left|\displaystyle\sum_{f_i} \frac{X^*\left(f_i - f_c\right)}{S_n\left(f_i - f_c\right)} \frac{X\left(f_i + f_c\right)}{S_n\left(f_i + f_c\right)} S_s^{2f_c}\left(f_i\right)^*\right|}{\displaystyle\sum_{f_i} \frac{|X\left(f_i\right)|^2}{S_n\left(f_i\right)} + \frac{\left|X\left(f_i + \frac{1}{T_s}\right)\right|^2}{S_n\left(f_i + \frac{1}{T_s}\right)}}. \tag{178}$$

Each of these ratio detectors possess a property not found in other detectors: they have a constant false alarm rate (CFAR) against changes in noise scale. The proof of this is very short, and is included here for completeness.

Theorem 2 (Cyclic Ratio Detectors are CFAR Detectors) *Each of the ratio detectors, shown in Eqns. (176), (177) and (178), have a constant false alarm rate (CFAR) against noise of a changing scale.*

Proof: *Since the proof for each of the three ratio detectors is essentially identical, only the first will be presented. To show that $y_{RATIO}(\boldsymbol{x})$ is CFAR, we examine the H_0 case when $\boldsymbol{x} = \boldsymbol{n}$. Under this case, let η be a detection test threshold chosen based upon the true probability distribution of $y_{RATIO}(\boldsymbol{n})$ such that*

$$P\left[y_{RATIO}(\boldsymbol{n}) > \eta\right] \;=\; P_{FA}. \tag{179}$$

Now, suppose the noise scale changes suddenly and $\sigma\boldsymbol{n}$ is received instead of $\boldsymbol{n}$. Examining $y_{RATIO}(\sigma\boldsymbol{n})$ shows that this statistic is invariant to this scale change,

$$
\begin{aligned}
y_{RATIO}(\sigma\boldsymbol{n}) &= \max_{\tau,\theta} \frac{\sigma\boldsymbol{n}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \sigma\boldsymbol{n}}{\sigma\boldsymbol{n}^\dagger \mathbf{R}_n^{-1} \sigma\boldsymbol{n}} \\
&= \max_{\tau,\theta} \frac{\boldsymbol{n}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{n}}{\boldsymbol{n}^\dagger \mathbf{R}_n^{-1} \boldsymbol{n}} \\
&= y_{RATIO}(\boldsymbol{n})
\end{aligned}
\tag{180}
$$

Given that $y_{RATIO}(\boldsymbol{n}) = y_{RATIO}(\sigma\boldsymbol{n})$, the false alarm rate is given by,

$$P\left[y_{RATIO}(\sigma\boldsymbol{n}) > \eta\right] = P\left[y_{RATIO}(\boldsymbol{n}) > \eta\right] = P_{FA}. \tag{181}$$

Thus, the false alarm rate remains constant across a changing noise scale, as required. Q.E.D.

Looking at these two new detectors, both the *multicycle ratio* and the *single cycle ratio* detectors, each of them uses a measurement "orthogonal to the signal," which is included in the total energy measurement in the denominator. This allows them to track changes in the background noise and detect only signals of interest. Two examples will help to illustrate this.

First, suppose a broadband signal of no interest was present during the observation interval but not during the training interval under which $\mathbf{R}_n$ was estimated. This broadband signal would raise the value in the numerator of both the single cycle and multicycle ratio detectors. Unlike the radiometric or cyclic feature detectors, however, the denominator would also increase, eliminating the effect of the rise in the numerator. This should render the detector immune to changes in broadband noise.

Consider, as a second example, an interferer having a nearly identical power spectral density to the signal of interest. In this case, the $\alpha = 0$ term of the multicycle detector would respond favorably to this signal, suggesting a signal of interest is present. This one increase could be large enough to create an alarm in any detector that used the $\alpha = 0$ term either by itself, or as one of several in a linear combination of terms. This is the common justification for throwing out the $\alpha = 0$ term while creating a signal selective detector. With a little foresight, one might recognize that the other cyclic terms, such as the $\alpha = \frac{1}{T_s}$ or $\alpha = 2f_c$ terms, would also measure a corresponding increase in their variance. This increase, coupled with the necessity of solving for τ and possibly θ, would cause detectors built from these terms, such as all common cyclostationary detectors, to alarm as well. The response of the cyclic ratio detector, however, would be tempered by a corresponding increase in the denominator, preventing such false alarms. This would provide the cyclic ratio detector a certain amount of immunity to burst interference.

Tests in the next chapter will focus on each of these scenarios in turn. As the foregoing discussion suggests, the cyclic ratio detectors will be shown to achieve

a signal selectivity that no other detector possesses. This fulfills Sonnenschein's prediction that, "the performance of detectors ... would be improved by this scheme since it would effectively remove the noise–level uncertainty" [60, p. 367].

This cyclic ratio detector is a fundamentally new type of detector. By its design it is much more resistant to burst interference than either energy detectors or other cyclostationary detectors. This exceptional resistance is demonstrated in Sec. 4.4.

3.5 Conclusions

This chapter introduced a new approach to burst signal processing based upon a new representation of digital communications signals in frequency. This new representation differed from other cyclostationary models in four respects. First, it accounted for the underlying redundancies within a communications signal's spectra, allowing the communications signal to be represented as a subspace of the received waveform. Second, by assuming that the data was Gaussian, a reasonable probability density function was applied to represent the data. This made it possible, later on, to apply classical statistical principles to the application areas of interest. Third, this model allows for the easy description of noise plus interference, making noise removal part of every initial algorithm development, rather than an afterthought. Finally, unlike previous models for cyclostationary signals, this model is quite appropriate for burst signals. Together, these differences make this model ideal for creating signal processing algorithms applicable to burst signals in colored noise.

This new approach was then applied to three separate application areas, resulting first in the development of well–known algorithms when the assumptions used were identical to those of the previous derivations. For example, Berger and Tufts' MMSE filter was shown to be a consequence of having a complex baseband. Then, when applied to TDOA estimation, this approach lead to the optimal stationary TDOA estimation filter. Likewise the optimal, locally most powerful, detector for detecting a cyclostationary signal in colored noise was shown to be the multicycle

detector presented by Gardner and Rostaing [17,51]. By arriving at these known solutions, each of these examples supports the contention that this is a valid approach for deriving new cyclostationary signal processing algorithms.

New algorithms were then developed, through the application of classical statistical principles, but this time under new assumptions. First among these were optimal single and multichannel demodulation filters. If truly optimal, as the next chapter will demonstrate, these filters demonstrate the validity of the estimate of the signal derived from this model. Not only were optimal filters derived, but by measuring the mean square error in an estimate, the mean square error at the output of a demodulator can be calculated under arbitrary noise and channel conditions. Second, when applied to TDOA estimation, the full form of a maximum likelihood TDOA estimator in colored noise was derived. Then, using this model, the first ever Cramér–Rao bounds were derived for cyclostationary TDOA estimation. Finally, when this model was applied to detection, new interference resistant detection algorithms emerged. All of these new results, from filtering algorithms to new signal selective detectors, are simple consequences of applying classical statistical principles to this fundamentally new model for cyclostationary signals.

What remains to be shown in the next chapter is the performance improvement that can be expected from using these algorithms and, equivalently, the performance loss from using suboptimal approximations. The performance improvements, in particular, will validate that these algorithms do indeed meet or exceed the performance of other stationary and cyclostationary algorithms when tested under conditions similar to those they were derived under. This will be demonstrated in the next chapter.

IV. Analysis by Simulation

Having developed a new approach to generating signal processing algorithms for digital communications in the last chapter, this chapter takes that development one step farther by validating the new methods derived from this framework in a simulated environment. Given that this new approach was founded on the assumption that the underlying signal had a Gaussian probability distribution in frequency, this assumption is first examined in detail before proceeding to the new algorithms. Once validated, the maximum likelihood estimates created from this probability distribution are then compared to other similar estimates, since these estimates will eventually be used in any detection algorithm. Thus the second and third section of this simulation chapter will examine the capability of the MMSE filters for generating an estimate of the data and then the two–sensor maximum likelihood TDOA estimator. Once these estimates have been shown to outperform all others tested, the final section examines the new detection algorithms derived under this model. Together, these simulations will demonstrate that the linear subspace approach to communications signal processing yields either complementary or superior results to those methods presented in Chapt. II.

4.1 Signal Model

Before looking at any of these new algorithms, however, we first validate the Gaussian assumption underlying all them. In particular, the fundamental assumption in Chapt. III was that $D\left(e^{j2\pi(f-f_c)T_s}\right)$ could be modeled as a multivariate Gaussian vector in frequency. Although this assumption was justified for large numbers of symbols by the Central Limit Theorem, it remains to be seen how well it applies to shorter signals. As this section will show, the assumption is reasonable in essentially all cases.

To show how well this assumption applies, we shall compare the true probability distribution of a very short QPSK signal in frequency compared to a Gaussian. While one might desire to compare probability density functions, these functions do not exist for discrete probability distributions such as those used in digital communication. Therefore, we instead compare the *Cumulative* Distribution Function (CDF) of $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ and its moments.

Prior to examining the CDF, however, some parameters need to be chosen. A extremely short signal, $N_s = 8$, was chosen to show how quickly this probability distribution converges to a Gaussian. Second, for simplicity, only the real portion of $D\left(e^{j2\pi(f-f_c)T_s}\right)$ will be examined. This makes sense since, for a QPSK signal, both real and imaginary portions have identical probability distributions. This leaves open the question of what frequency values to use in this comparison.

To resolve this issue, Fig. 8 shows the CDF of $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ as a function of radian frequency, $\omega = 2\pi\left(f - f_c\right)T_s$, and symbols, x. The radian frequency is shown from $\omega = 0$ to $\omega = \frac{\pi}{4}$ only since the rest of the radian frequency band is symmetric—repeating this same pattern. The second axis shows the range of $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ in symbols. Vertically, this plot shows the probability that $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ is less than x symbols at some radian frequency ω. From Fig. 8, we see that this CDF converges quickly to a smooth function for most frequencies. Only the middle, $\omega = \frac{\pi}{8}$, and edge, $\omega = 0$ and $\omega = \frac{\pi}{4}$, frequencies do not to converge as quickly to this smooth function. Of these two, convergence is worst at the edge.

What is not necessarily obvious from Fig. 8 is that this smooth function describes a Gaussian probability distribution. To see this, consider a slice of this the CDF of $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ taken from somewhere in the middle, say $\omega = \frac{\pi}{10}$, and compare it to the CDF of a true Gaussian having the same mean and variance (Fig. 9). From this vantage point, the two CDFs are nearly on top of each other. This shows that the smooth function, which the overall CDF appeared to converge to, is indeed a Gaussian CDF.
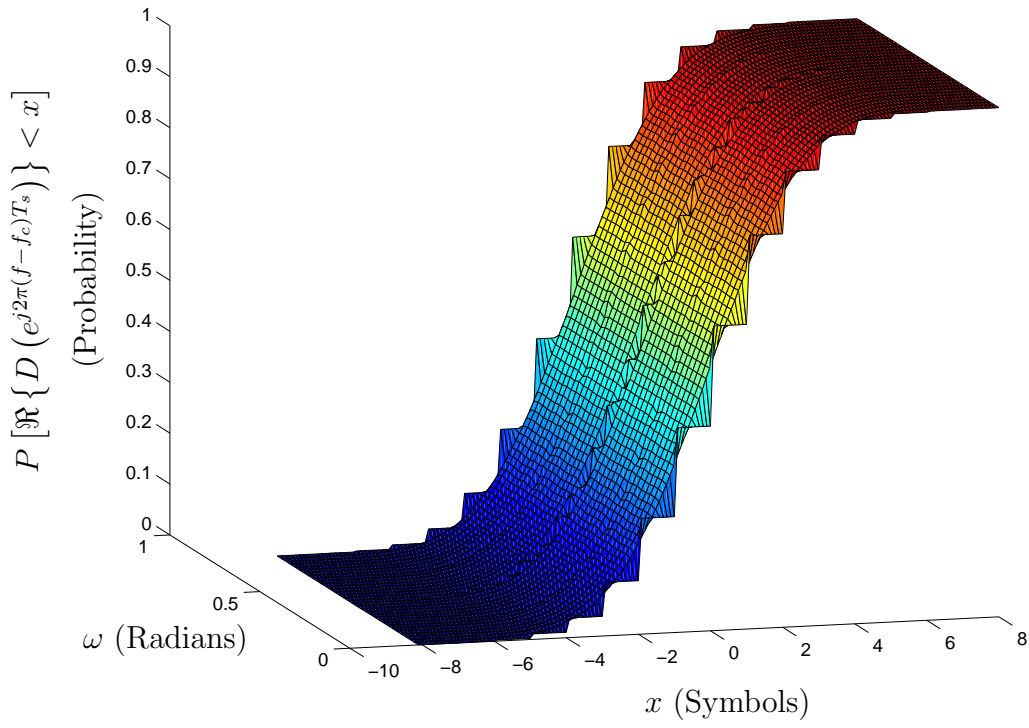
Figure 8.   Cumulative Distribution Function
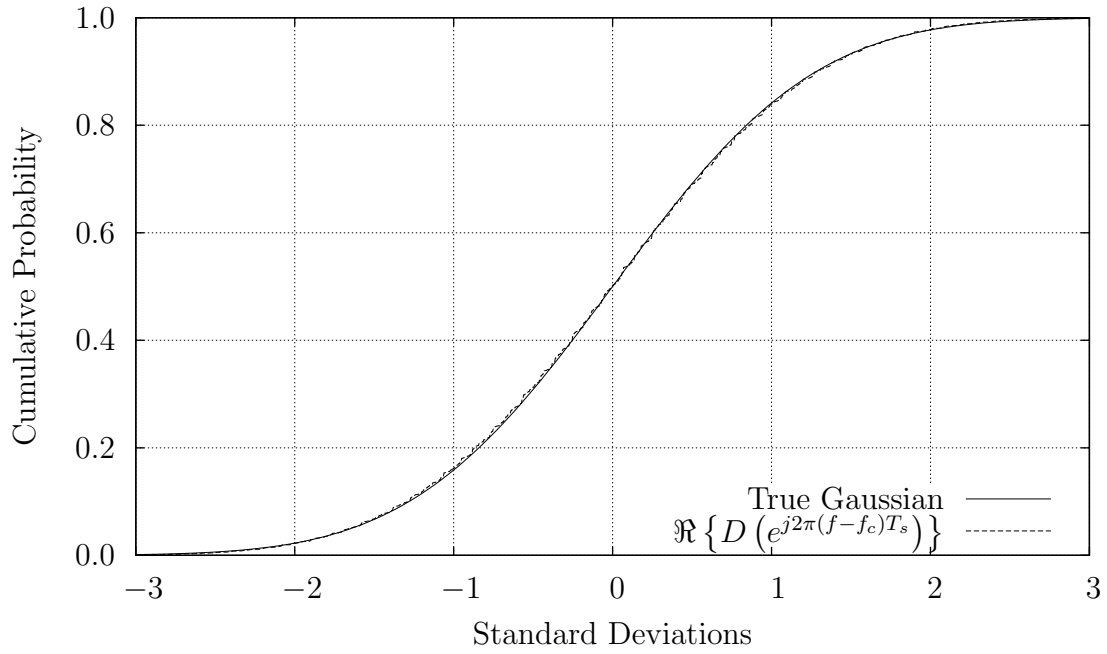
Figure 9.   CDF of $D\left(e^{j2\pi(f-f_c)T_s}\right)$ when $N_s = 8$ and $f = f_c + \frac{1}{20T_s}$

108

1.5
1.6
1.7
1.8
1.9
−4
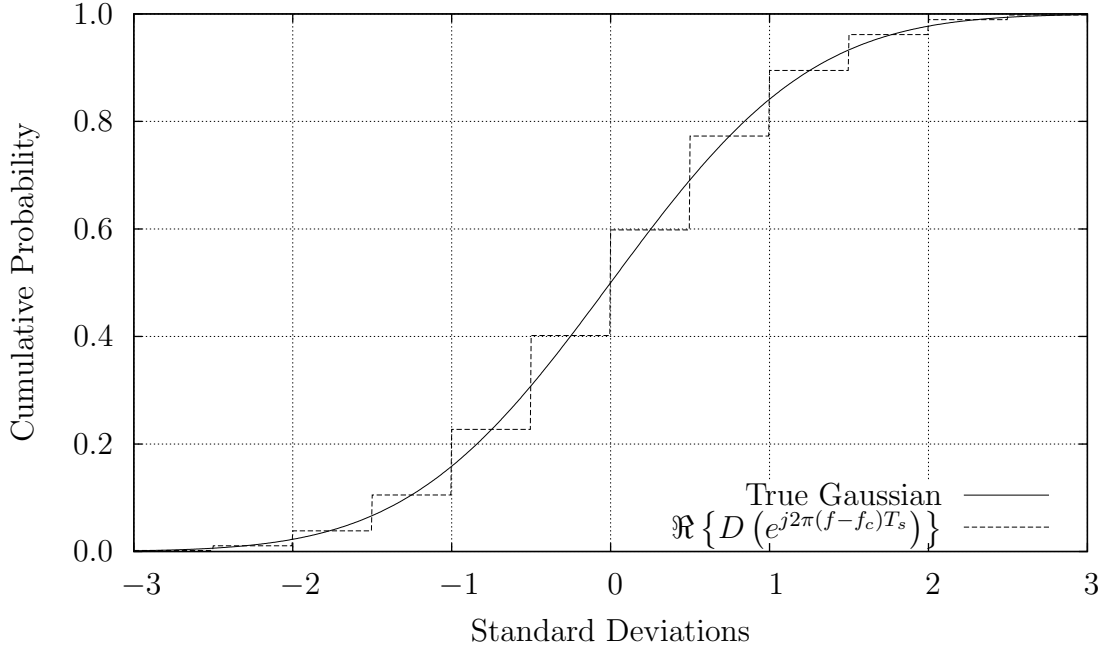
Figure 10.   CDF of $D\left(e^{j2\pi(f-f_c)T_s}\right)$ when $N_s = 8$ and $f = f_c$.

Yet this example only shows the best case convergence. To examine a worst case convergence, we compare a Gaussian CDF with the CDF for $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ at the edge ($\omega = 0$) in Fig. 10. In this case, the true CDF approximates a Gaussian with step functions. While this step function behavior holds for all values of N_s, the steps do get shorter as N_s increases. This is illustrated by Fig. 11, which shows the same two CDF functions, only this time for a signal having $N_s = 256$ symbols instead of $N_s = 8$.

From these three figures, we conclude that the true probability distribution of $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ is roughly Gaussian for even short bursts. When the approximation is poor, such as in Fig. 10, it is at least as good as a step function converging to a Gaussian CDF. When the approximation is good, such as in Fig. 9, it appears to match very well. In all of these cases the approximation is quite reasonable.

The final way of demonstrating that true probability distribution is approximately Gaussian is to consider the moments or $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$. If this approx-

109

1.5
1.6
1.7
1.8
1.9
−4
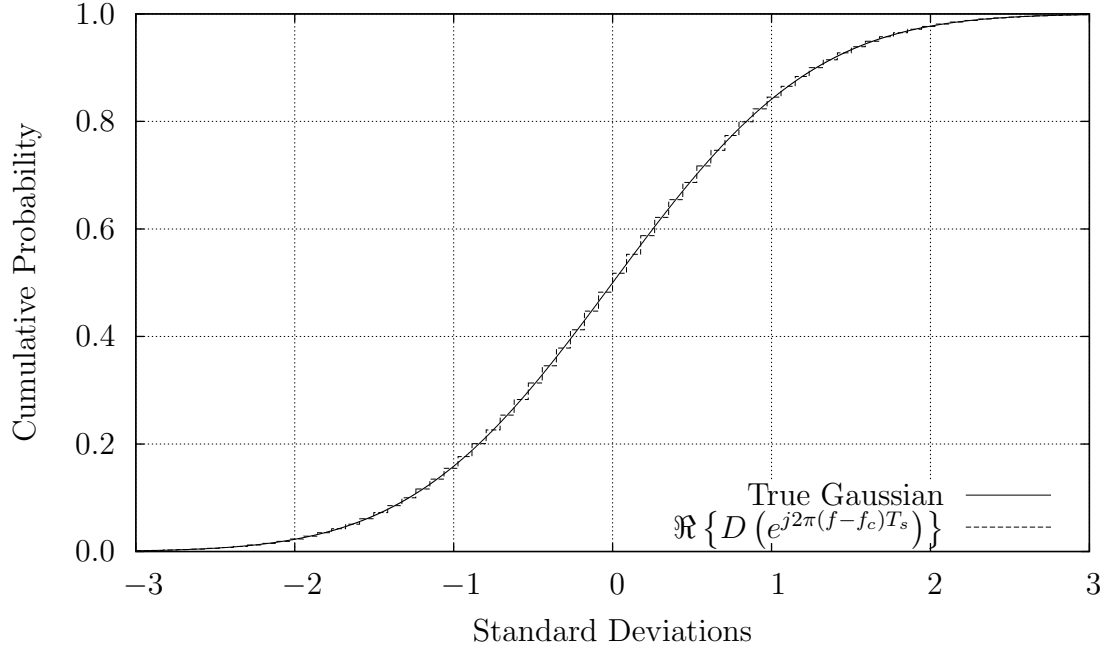
Figure 11.   CDF of $D\left(e^{j2\pi(f-f_c)T_s}\right)$ when $N_s = 256$ and $f = f_c$.

imation is valid, one would expect the moments to resemble those of a Gaussian distribution. Therefore, Fig. 12 examines the relative error between the fourth and sixth moments of $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ for $\omega = 0$ and the fourth and sixth moments of a true Gaussian. As one might expect from the central limit theorem, the actual moments of $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ converge to those of a Gaussian as N_s increases. In addition, the fact that the slope of this convergence is -1 shows that this convergence is linear as the number of symbols increases.

These two examples demonstrate that approximating the probability distribution of $\Re\left\{D\left(e^{j2\pi(f-f_c)T_s}\right)\right\}$ as a Gaussian, while not perfect, is at least reasonable. From Sec. 3.1.2, we expected this approximation to be valid as N_s grew large. Here, we saw that it did in fact converge as N_s grew arbitrarily large. What was not necessarily expected from Sec. 3.1.2 was that, for really short bursts, the true probability distribution still appeared to be roughly Gaussian. Even in the worst case, when $\omega = 0$, the true probability distribution corresponds to the well–known binomial dis-
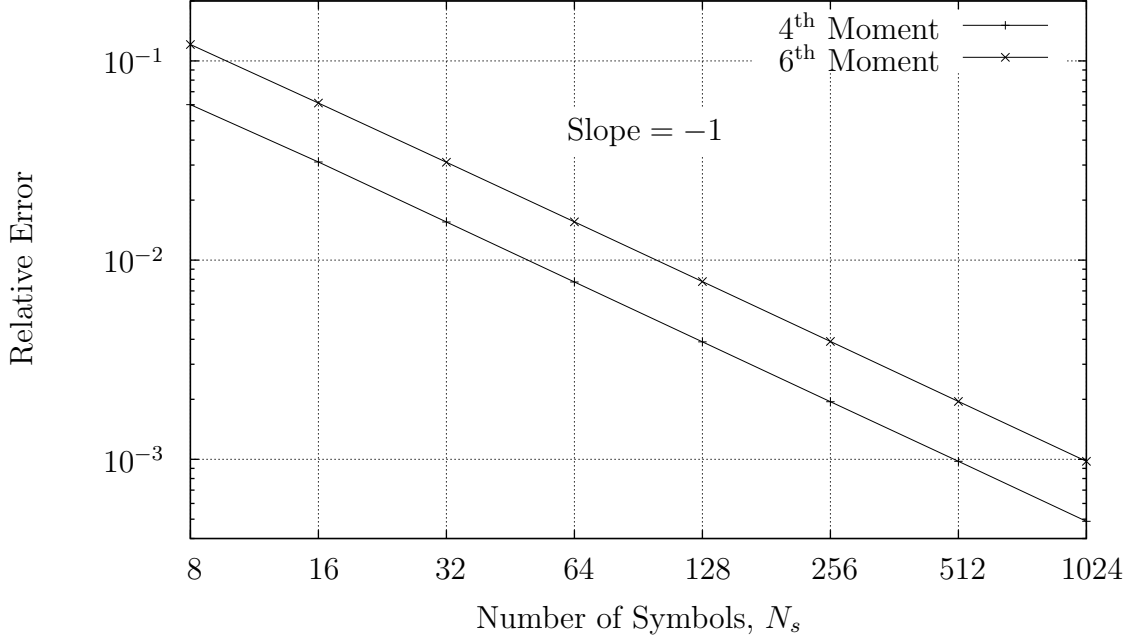
Figure 12.   Convergence of the Moments

tribution, which is commonly approximated by a Gaussian [7]. For all other values of ω, the approximation is much better. This justifies the assumption, used throughout this research, that the probability distribution of $D\left(e^{j2\pi(f-f_c)T_s}\right)$ is a Gaussian.

4.2   BPSK Filtering

Having demonstrated the validity of the Gaussian approximation for even the shortest burst signals, the next step is to look at the results of the subspace approach, all of which were founded upon this assumption, and to evaluate them through simulation. As before, the first application area needs to be filtering because appropriate filters become the estimators used in every subsequent application area. That is to say, if the MMSE filters developed in Sec. 3.2 fail to perform well in simulation, then there is no reason to expect good performance from any other algorithm derived from this linear subspace approach. Instead of failure, however, this section will demonstrate that the new filters for BPSK signals outperform every other filter tested here, from matched filters to Berger and Tufts MMSE filter.

111

One by one, each of the three new demodulation related algorithms derived in Sec. 3.2 will be demonstrated in this section. First, the single sensor BPSK MMSE filter will be shown to outperform every other filter it is tested against in a severe interference environment. Next, the performance prediction methods will be applied to those same filters. The versatility of this method is demonstrated in both the severe interference environment from the first test as well as a strong multipath environment. In each case, this method correctly predicts the MSE degradation resulting from not compensating properly for the environment. Finally, the multi–sensor filters will be demonstrated on a two sensor example with both wideband and narrowband interference. As with the single sensor results, this last subsection will demonstrate that the optimal two sensor filter outperforms all single sensor filters. Together, all three of these tests validate both the form of the optimal data estimator as well as the predicted loss associated from less than optimal estimation.

4.2.1 BPSK Minimum Mean Square Error (MMSE) Filters. The development of the single sensor BPSK MMSE filter introduced the hypothesis that this new filter will outperform all other linear filters, in terms of MSE, when demodulating BPSK signals. Demonstrating this optimality, however, requires generating a signal in a colored noise environment, and demodulating it with several potential filters of interest. Once demodulated, comparing the resulting symbol estimates with the original symbols yields either a mean square error metric, for which this method should be optimal, or a bit error rate metric, which will tell more of the capability of this method. This section, therefore, starts off with a description of the approach used to estimate these two parameters. Then, since all tests of this type are highly dependent upon both the signal and the interference environment, the next step will be to describe the parameters chosen for the signal followed by the background noise and interference. Once these have been described, simulation results can be presented that demonstrate the capability of this filter. As a last test, the hypothesis that an adaptive linear equalizer converges to the the equalizer in

112
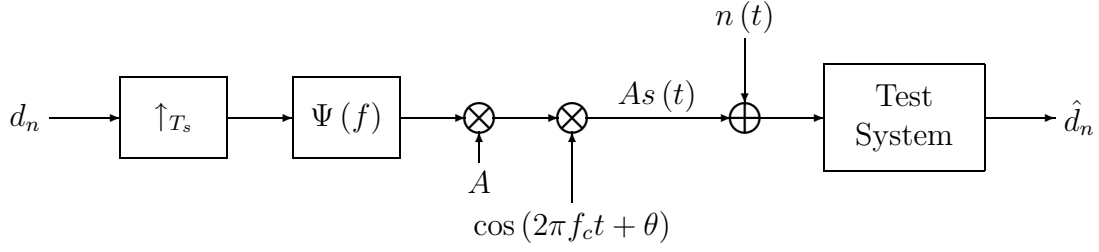
Figure 13. Filtering Test Setup

Eqn. (80) is tested. Such equalizers are well known and proofs of their convergence to the MMSE solution are readily available [30]. This last comparison will do more than validate this filter, it will in fact prove that it truly does outperform *all other* BPSK demodulation filters in this environment.

Since the goal will be eventually to compare theoretical results with results achieved via simulation, a well–calibrated simulation will be required to run these tests. Such a simulation, outlined in Fig. 13, may be created digitally by constructing a simulated signal, $As(t)$, from a random symbol sequence, d_n, and then by adding the result to a simulated noise sequence. This sequence will then model what a digital receiver might measure coming off of its antenna. The next step is to run the received signal through the system under test, whether it be the baseband demodulator shown in Fig. 5 on page 64 for Berger and Tufts' filter, or the bandpass demodulator shown in Fig. 6 on page 65. The estimated symbols at the output of this system, $\hat{d}_n$, will then be compared with the original symbols which were used to generate the signal. From this comparison, an estimate may be generated for the MSE in the demodulator,

$$\text{MSE} \approx \frac{1}{N_s} \sum_{n=0}^{N_s-1} \left(\hat{d}_n - d_n\right)^2, \tag{182}$$

113

and similarly for the BER at the output of the demodulator,

$$\text{BER} \quad \approx \quad \frac{1}{N_s} \sum_{n=0}^{N_s-1} \begin{cases} 0 & \hat{d}_n d_n \geq 0 \\ 1 & \hat{d}_n d_n < 0. \end{cases} \qquad (183)$$

This process is then repeated for multiple values of the signal gain, A. All that remains is to describe how the noise and signal sequences are created and placed into this test sequence.

The first step in this process, creating the signal, follows the process shown in Fig. 13 quite literally. This process involved first generating a random data sequence, $d_n \in \{\pm 1\}$, and then upsampling this data sequence by T_s, which was chosen to be 10 samples thus yielding a symbol rate of 0.1 cycles per sample (CPS). The resulting impulses were then smoothed by the pulse shaping filter, $\Psi(f)$, which was chosen to be the Nyquist pulse defined in App. A, with the exception that it was tapered to 32 symbols in length via a Hanning window [47]. Further, to simplify energy measurement, this Nyquist pulse was normalized so that the energy transmitted per bit was simply $E_b = A^2$. The signal was then multiplied by an adjustable gain, A, and a carrier, $\cos(2\pi f_c t + \theta)$. The carrier frequency for these simulations was chosen to be 0.2 CPS. Together, these choices were made to place the signal in the center of the normalized frequency band and to minimize any aliasing effects from sidelobes. Finally, τ and θ were allowed to be known exactly rather than estimated, following the assumptions in Chapt. III. These parameters are summarized in Table 3.

The noise, on the other hand, was chosen to be a combination of both a white noise background and a colored noise spike, as shown in Fig. 14. Although other noise environments, such as the two–sensor environment that will be used in Sec. 4.2.3, will result in better performance of this filter over the others tested here, this environment was chosen because it clearly illustrates the operation of the filter. The noise spike itself was obtained by passing a second white noise sequence through a simple two

114

Table 3. Signal Parameters for the Single Channel Filter Test

Signal Parameter	Parameter Value
Modulation Type	BPSK
Carrier Frequency, f_c	0.2 CPS
Symbol Rate, $\frac{1}{T_s}$	0.1 CPS
Pulse Shape, $\psi(t)$	Nyquist
Carrier Phase, θ	Known
Symbol Epoch, τ	Known

pole, real, IIR filter. The filter was designed to place the spike just to the right of the carrier of the BPSK signal, creating an asymmetric noise profile. The reason for this asymmetry should be clear: had a symmetric noise profile been chosen then the true MMSE filter would have been no different from Berger and Tufts' filter. Finally, the height of the spike was chosen to dwarf everything else in the environment. As the following discussion will demonstrate, this noise spike makes it easy to see and follow the differences between the BPSK MMSE filter and the other filters of interest.

Adding the signal into this noise environment results in a received power spectral density such as the one shown in Fig. 14. This figure clearly shows that the noise spike, for small to moderate signal energies, totally dominates the signal as desired.

Each of the four filters, shown in Fig. 15, were tested and compared in this environment. These filters are the matched filters (MF) for white and for colored noise, Berger and Tufts' baseband filter, and the BPSK MMSE filter developed in Sec. 3.2.1. Starting with the simplest, the transfer function of the matched filter for white noise has an identical shape as the underlying pulse shape in frequency.[1] Unlike all of the other filters tested, the matched filter for white noise does nothing to compensate for the interference spike. The matched filter for colored noise, on

---

[1]Normally, the frequency response for this filter would be the conjugate of the pulse shape, but since the Nyquist pulse defined in Appendix A is symmetric in time about zero, the corresponding matched filter frequency response is entirely real.

0.16
0.18

0.22
0.24

-0.26
-0.28
-0.10
-0.05
0.35
0.40
0.45
0.50
0.52
0.54
160
0.40
0.18

0.22
0.24

0.26
0.28
−20

0.35
0.40
0.45
0.50
−3000
−2000
−1000

120

100

80

60

40

20

0

Power / Normalized Frequency

- Signal PSD, $E_b = 15$ dB
- Noise PSD
- Received PSD

Good quality
signal at
0.185 CPS.

Poor quality
signal at
0.215 CPS.

0.00   0.05   0.10   0.15   0.20   0.25   0.30

Normalized Frequency (Cycles / Sample)
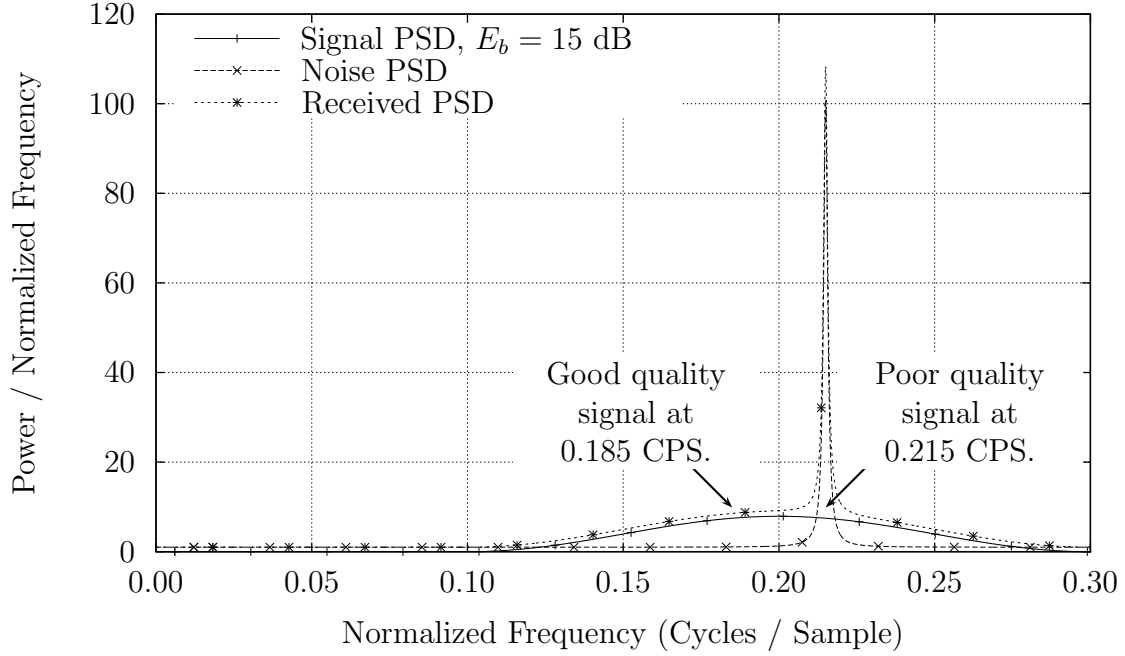
Figure 14.   Noise and Signal PSDs, when the signal $E_b = 15$ dB

5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0

5500
6000
7000

Filter Gain, $H(f)$, (Unitless)

BPSK MMSE
Berger and Tufts' MMSE
Colored Noise MF
White Noise MF

Baseband notch
to remove down-
converted inter-
ference spike

Notch to remove in-
terference spike

0.00   0.05   0.10   0.15   0.20   0.25   0.30

Normalized Frequency (Cycles / Sample)
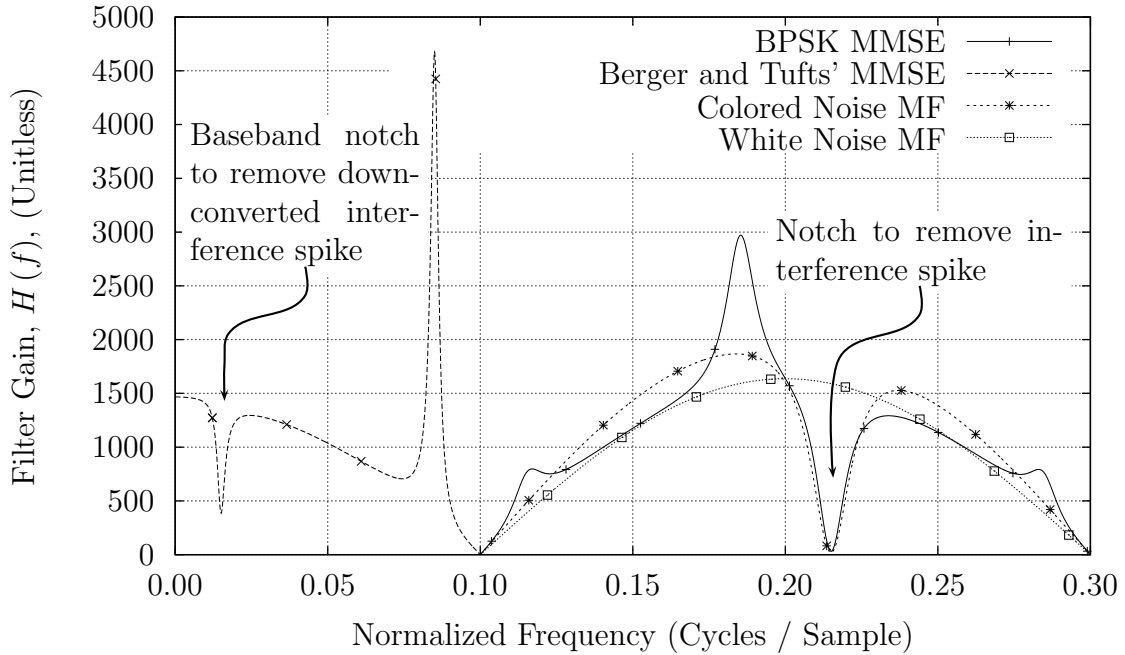
Figure 15.   Transfer Functions for Several "Optimal" Filters. These filters have been
normalized to each produce unity gain, $\gamma = 1$.
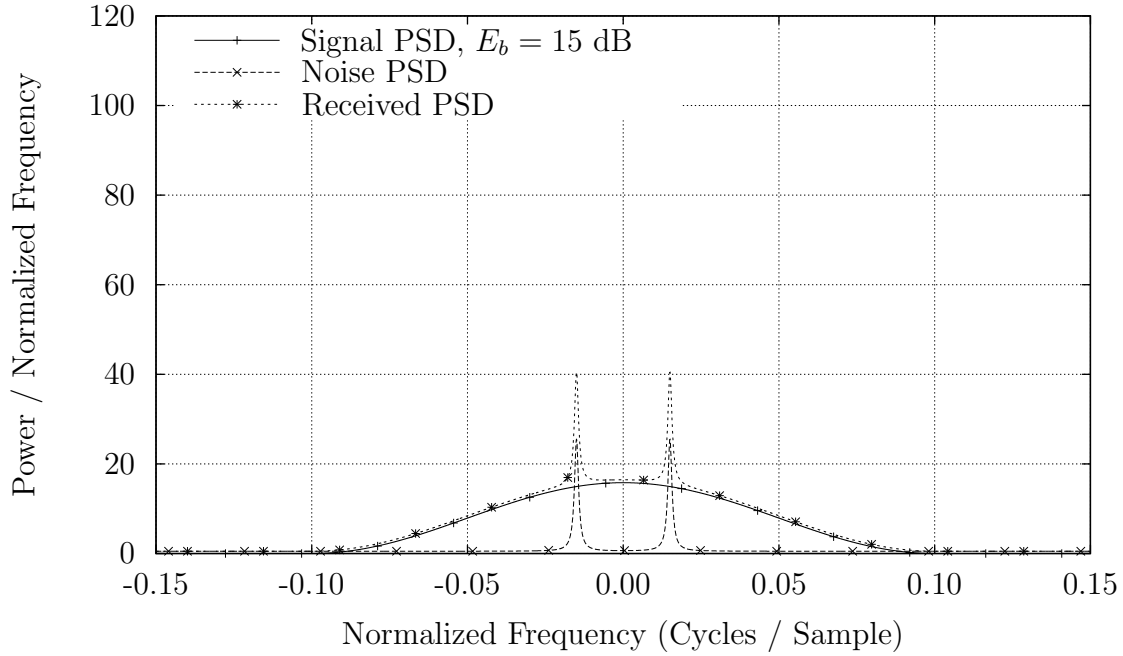
Figure 16.　Unfiltered Baseband PSD

the other hand, notches out the interference as one might expect, but does nothing to compensate for any distortion that might be introduced by such a notch. Finally, the BPSK MMSE filter not only notches out the interference, but also boosts the signal strength at the three other frequency locations corresponding to the same underlying data information as the notch.

Berger and Tufts' filter, however, needs a little more description. Since this is a baseband filter, it has no response to the frequencies between 0.1 and 0.3 CPS like the other filters in Fig. 15. Instead, this filter is not applied until after the signal has been downconverted.[2] As a result, its response in Fig. 15 lies between 0 and 0.1 CPS.[3] This process, however, also downconverts the noise spike, placing it at ± 0.015 CPS and creating the PSD shown in Fig. 16. This downconversion merges the high SNR

---

[2]The system diagram for this filter is shown in Fig. 5 on page 64 for reference.

[3]Since Berger and Tufts' filter is a real, as opposed to complex, filter, its response for negative frequencies is the complex conjugate of its response for positive frequencies. In this case, it is just symmetric about the zero frequency.
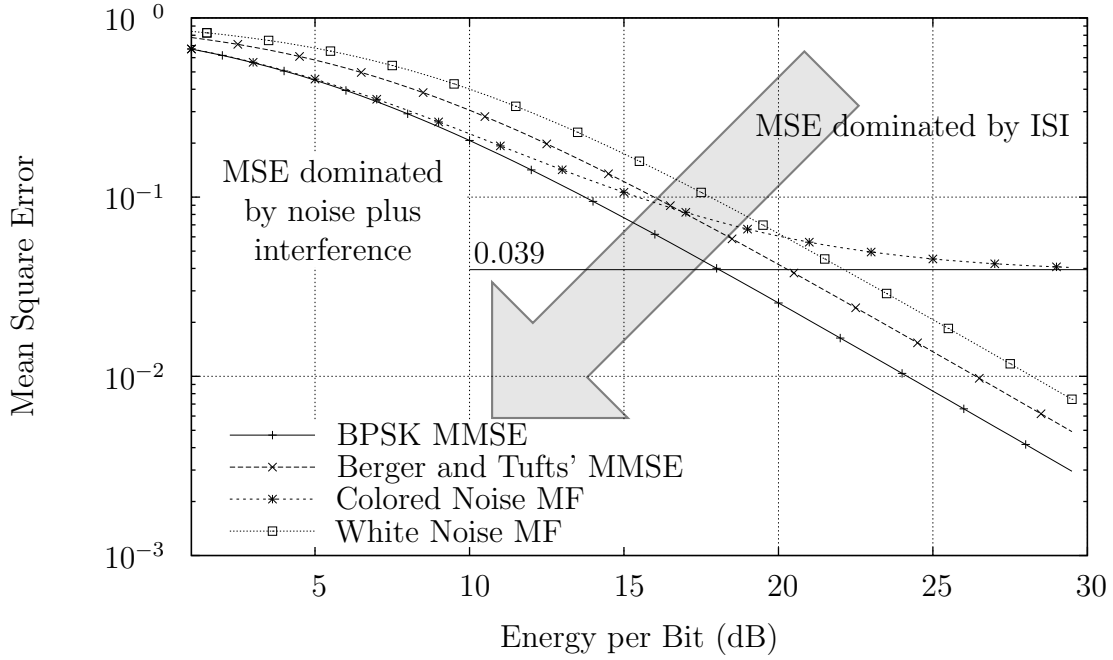
Figure 17.   Mean Square Error Comparison

signal, shown in Fig. 14, at 0.185 CPS with the poor signal at 0.215 CPS prior to removing the noise. Only after these two components are merged, creating the PSD shown in Fig. 16, is the noise spike notched out. Then, compensating for this notch requires boosting the weakest redundant component of the signal at ± 0.085 CPS (the tallest spike in Fig. 15). This whole process prevents the effective application of the strong signal component initially found at 0.185 CPS. As a result, one might expect Berger and Tufts' filter to perform worse than the matched filter for colored noise which notches out this spike prior to downconversion, but better than a matched filter for white noise which does nothing to remove this interference.

To measure the performance of these filters, we first measure the MSE in a simulation environment since the BPSK MMSE filter was designed to minimize MSE, just like Berger and Tufts' filter. The MSE produced by each of the four filters from Fig. 15 is shown in Fig. 17 as a function of the energy per bit, $E_b = A^2$. From this figure, it is certainly plain that the optimal filter for low SNR conditions is the

matched filter for colored noise. The MMSE filter matches this filter in performance simply because it converges to the matched filter as the signal strength, A, goes to zero. Then, as predicted, Berger and Tufts' baseband filter did not perform as well. What may have been unexpected is the performance of the matched filter for colored noise as the signal strength increased. This matched filter asymptotically approaches a lower bound in performance that the other filters are not subject to.

The reason for this poor performance is actually quite straight forward. All of the mean square error, when no noise is present, must be due to Intersymbol Interference (ISI). Why? Consider, the Matched Filter (MF) for white noise does not suffer from ISI since the Nyquist pulse was designed to have no ISI following a matched filter for white noise. Therefore, as signal strength increases, the MSE decreases for this filter. Berger and Tufts' MMSE filter, together with the BPSK MMSE filter, each include an equalizer to remove any ISI induced by the overwhitener. The matched filter for colored noise, however, has no such compensation for ISI. Where did the ISI come from? It must have been from the overwhitener—the only difference between the matched filter for white noise and the matched filter for colored noise.

To confirm the hypothesis that the poor performance of the matched filter for colored noise was due to ISI, the ISI was calculated for each of the matched filters and for the BPSK MMSE filter when $E_b = 30$ dB. Fig. 18 shows the log of the absolute value of the contributions, from other symbols, to the current symbol of interest. As expected, the matched filter for white noise has a minimal amount of intersymbol interference associated with it. As hypothesized, the matched filter for colored noise has the highest intersymbol interference. Further, if you calculate the MSE expected from this ISI interference alone,[4]

$$\xi_{\text{ISI}}^2 \quad = \quad \min_{\kappa} \frac{2T_s}{N_s} \int_0^{\frac{1}{2T_s}} \xi_i^2 \left(e^{j2\pi f T_s} \right) df \tag{184}$$

---

[4]This formula follows from Eqn. 86, save that the MSE due to noise and other interference has been removed.
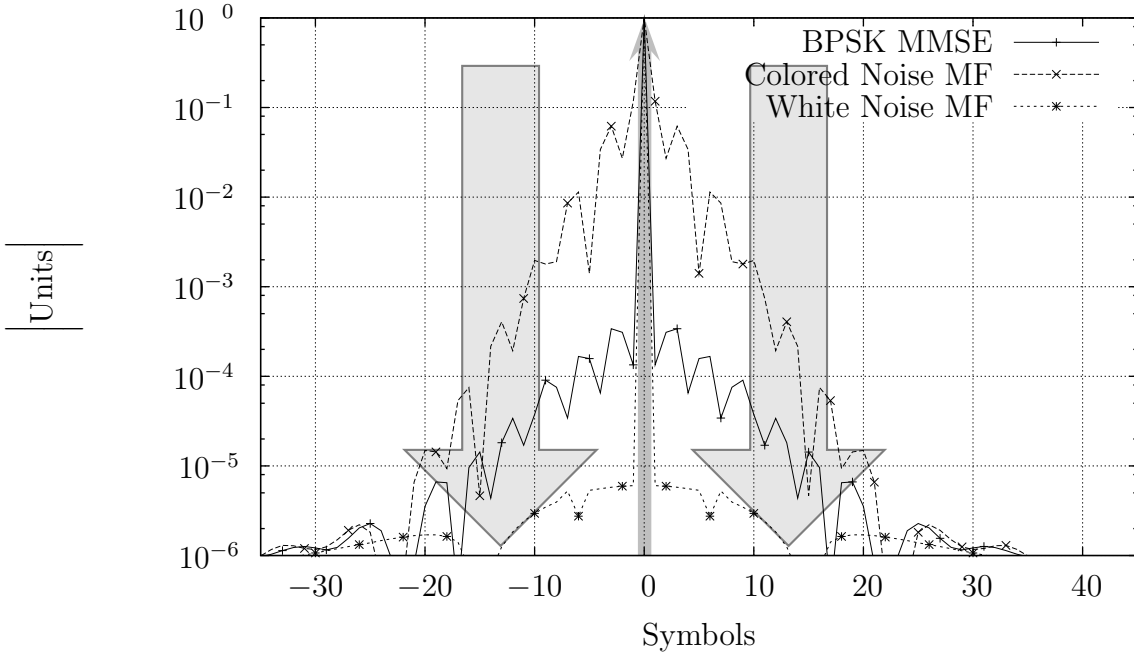
Figure 18.   Log Intersymbol Interference Function, $E_b = 30$ dB

you get 0.039 Units^2. Looking at the 0.039 Units^2 line in Fig. 17 confirms that this is indeed the source of the poor performance found in the matched filter for colored noise.

A much more meaningful metric, however, is the Bit Error Rate (BER). Unlike MSE, the BPSK MMSE is not guaranteed to achieve a minimum BER among all other filters. Yet, looking at the BER performance shown in Fig. 19, we see that the BER performance of these filters is very similar to their MSE performance. The only notable exception is the matched filter for colored noise again. This filter doesn't quite perform as poorly compared to the other filters in terms of BER as it did in terms of MSE. This is an artifact of the non–Gaussian nature of ISI which will be discussed in more detail in Sec. 4.2.2 on performance prediction.
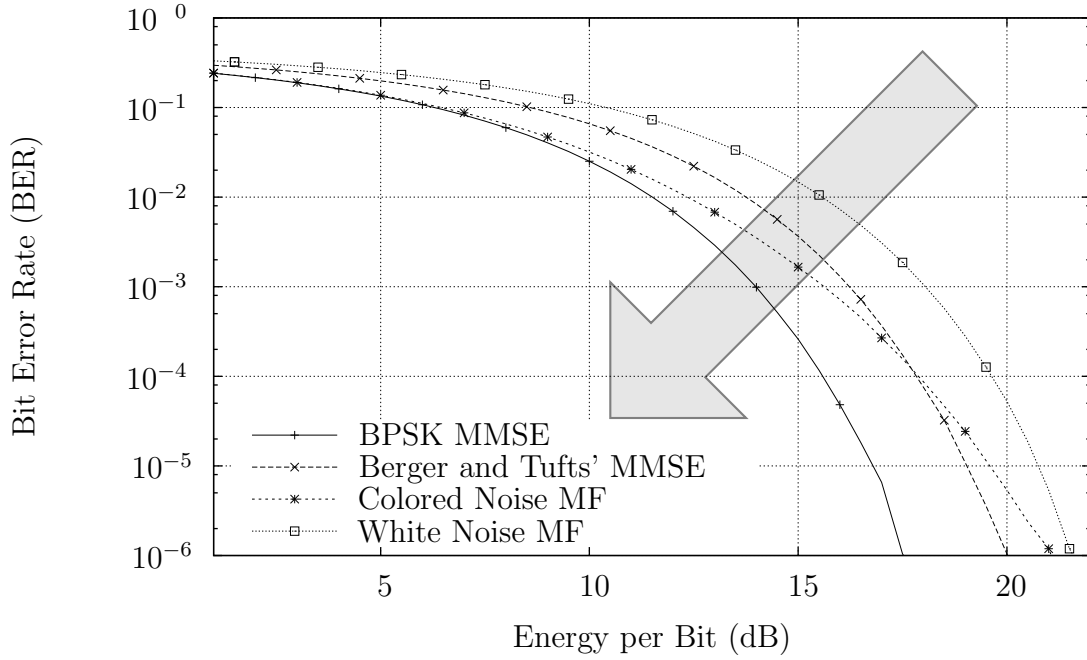
Figure 19.   Bit Error Rate (BER) Comparison

The final test of this filter is to compare it against a decision–directed adaptive linear equalizer.[5] The reason for this is to prove that this filter truly does achieve the minimum MSE among all other filters–including those not tested here. Here's why: According to [12], the optimal filter is always a matched filter for colored noise followed by an equalizer. Then, according to [30], adaptive equalizers are known to achieve the minimum MSE solution. Therefore, we test a demodulator formed by a matched filter for colored noise followed by an adaptive equalizer, shown in Fig. 20. The actual structure and implementation of this equalizer is described in [30], in the sections on "Channel Equalization" and then on the "Least–Mean–Square Algorithm," which was used to provide the adaptation. Finally, after this equalizer has been given a chance to converge, if the resulting equalizer is identical

---

[5]This equalizer is not to be confused with a non–linear decision–feedback equalizer. Such equalizers were not tested in this research.

121
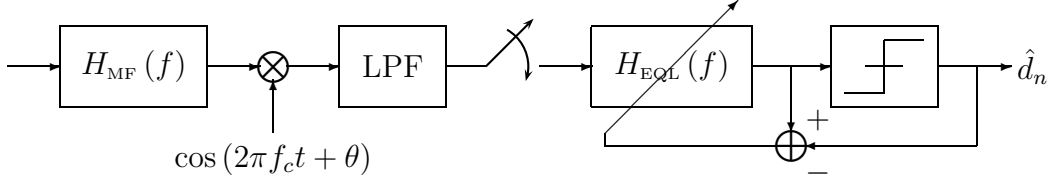
Figure 20.   A Decision–Directed, Linear, Adaptive Equalizer

to the predicted equalizer derived in Sec. 3.2.1,

$$
H_{\text{EQ}}\left(f\right) \;=\; \cfrac{1}{1 + \frac{A^2}{4}\frac{|\Psi(f-f_c)|^2}{T_s S_n(f)} + \frac{A^2}{4}\frac{\left|\Psi\left(f_c-f-\frac{1}{T_s}\right)\right|^2}{T_s S_n\left(2f_c-f-\frac{1}{T_s}\right)} + \frac{A^2}{4}\frac{|\Psi(f_c-f)|^2}{T_s S_n(2f_c-f)} + \frac{A^2}{4}\frac{\left|\Psi\left(f+\frac{1}{T_s}-f_c\right)\right|^2}{T_s S_n\left(f+\frac{1}{T_s}\right)}},
$$

(185)

then we say that the BPSK MMSE filter truly achieves the minimum MSE among all other filters in this environment.

After applying this filter to a BPSK signal of 6.4 million symbols, the filter had converged enough to plot Fig. 21. This figure shows two lines for the adaptive equalizer's response. The first is an upper error bar and the second is a lower error bar. The actual response, given complete convergence from an infinite length signal, would lie somewhere between these two error bars. What is important to notice from this figure, however, is that the predicted equalizer response lies right between the error bars of the adaptive equalizer's response. This supports the conclusion that the BPSK MMSE filter does indeed predict the necessary equalizer to achieve MMSE performance.

This proof, together with the first simulation, demonstrates that the BPSK MMSE filter truly achieves MMSE performance. The second test demonstrated that, although MSE is not equivalent to BER, the MMSE filter also had the minimum BER among all of the filters tested here. None of the other filters tested, whether they were matched filters or Berger and Tufts' baseband MMSE filter, performed better than the BPSK MMSE filter. This means that, at least for one sensor, the

122
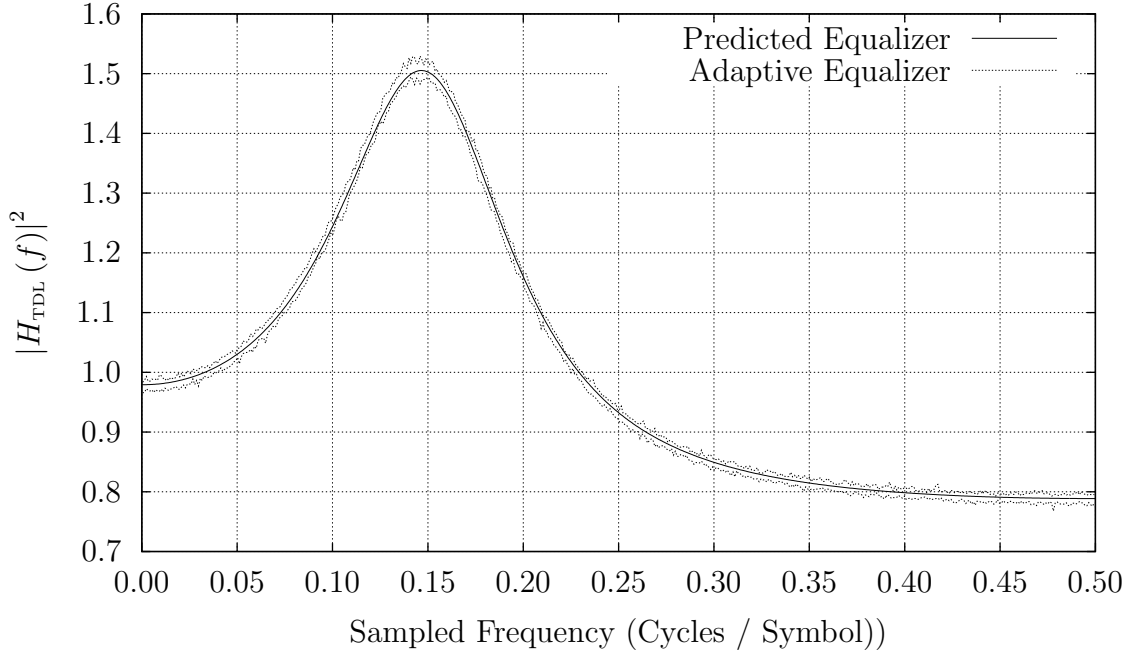
Figure 21.   Predicted verses Estimated Equalizer Response. The predicted response function goes right through the center of the estimated response function.

MMSE filter outperformed all of the other filters tested. This validates the proof in Sec. 3.2.1 that this filter has the lowest MSE among all filters, and lends credence to every other result following this section, all of which are founded upon this estimate.

   4.2.2   *Predicting BPSK Demodulator Performance.*   Having shown that a MMSE filter could be applied to achieve a lower MSE than all other linear filters, this section continues the MSE examination to demonstrate how the MSE at the output of an arbitrary filter can be calculated. Two tests are presented to demonstrate this concept. The first test demonstrates how the previous results could have been predicted using the MSE performance prediction methods developed in Sec. 3.2.2. The second test demonstrates the versatility of this method by examining the same prediction capability in a multipath environment.

   For the first test, the conditions were chosen to be identical to the filter performance tests in the previous section. A large noise spike was present, and the

123
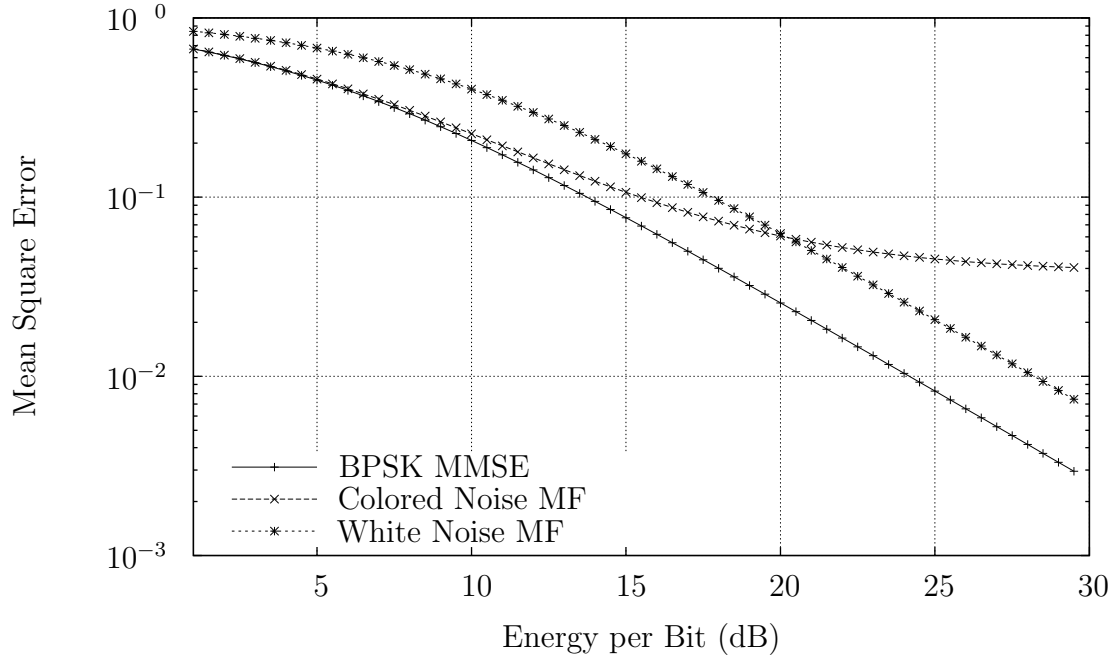
Figure 22.   Predicted Mean Square Error. The lines are the predictions, the points are arrived at via simulation.

demodulator was perfectly synchronized to the signal. Likewise the filters applied to the signal are identical save that Berger and Tufts' filter was not tested.[6]

As developed, these performance prediction formulas should predict the MSE in any BPSK demodulator exactly. To test this hypothesis, Fig. 22 shows the MSE measured at the output of the demodulator. Lines shown in this figure result from predictions made using Eqn. 88, while the points are the result of simulations. In every case, the prediction matches the simulation. This confirms that the prediction methods do in fact predict MSE as designed.

A more important metric is the BER. No claim was made regarding the prediction capability of these formulas for predicting BER, save that they would be accurate when the error disturbance was Gaussian. Since this is a more practical

---

[6]While nothing prevented Berger and Tufts' filter from being tested, doing so would have tested a baseband performance prediction formula that was slightly different from the formula derived in Sec. 3.2.2. Performance for Berger and Tufts' filter can still be predicted according to the Eqn. (24) within [2].
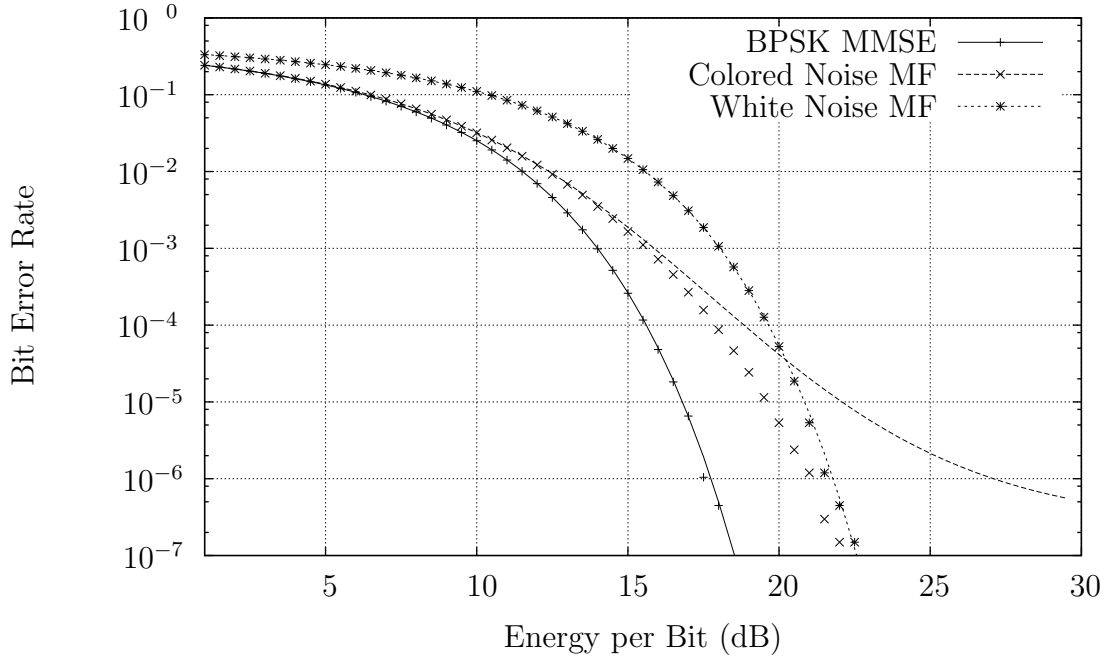
Figure 23. Predicted Bit Error Rate. The lines are the predictions, the points are arrived at via simulation.

metric in system design, we perform a test of the BER here. Thus Fig. 23 shows the BER resulting from the prediction method presented in Sec. 3.2.2, and in particular in Eqn. (92) on page 71, compared to the BER measured in practice. As with the previous figure, the lines on this figure plot the predicted performance while the points demonstrate simulated performance. This figure shows that, when the BER was the result of a (primarily) Gaussian disturbance, such as for the BPSK MMSE filter and the matched filter for white noise, the prediction matches. When the BER was primarily the result of an ISI type of disturbance, as opposed to a Gaussian one, the BER prediction was somewhat off. This highlights the assumption underlying this BER prediction method, that it is *only valid for Gaussian disturbances*. Other methods, such as those discussed in [34], are necessary for calculating the BER in severe ISI.

Next, in order to demonstrate the utility of this method, the experiment was repeated under severe multipath conditions. Since the performance prediction method

−0.4
−0.3
−0.2
−0.1

1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9

Power / Normalized Frequency

With Multipath Interference
No Interference

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

0.10 0.12 0.14 0.16 0.18 0.20 0.22 0.24 0.26 0.28 0.30

Normalized Frequency (Cycles / Sample)
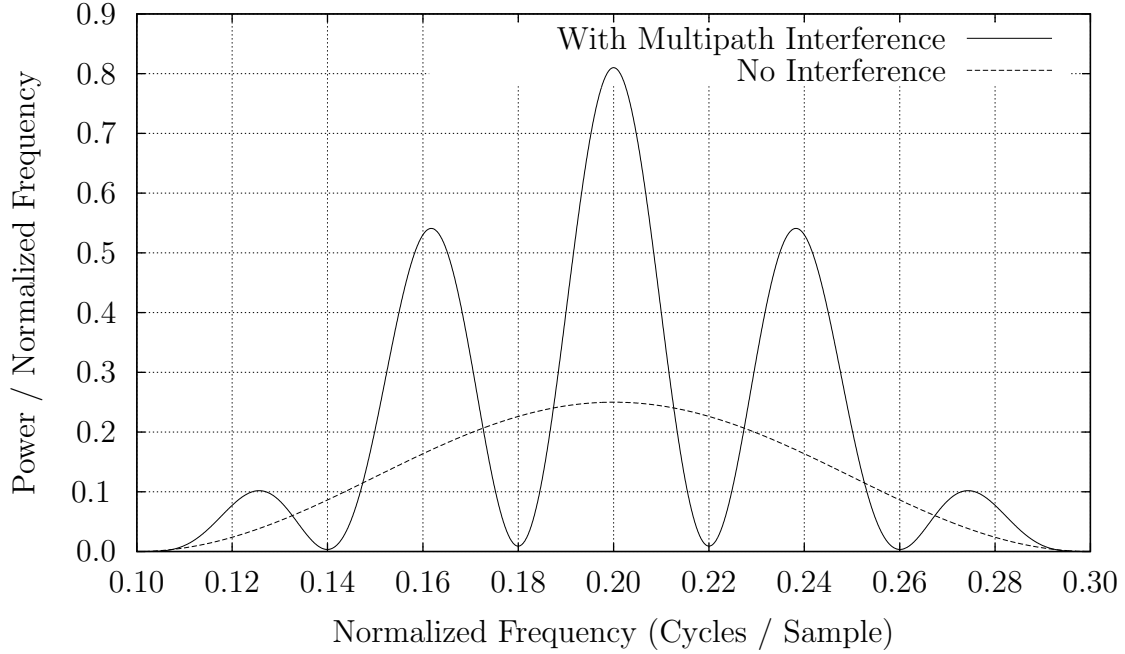
Figure 24.   Distortion Caused by Multipath

should work regardless of the multipath conditions, a simple yet severe environment
was created to distort the signal. In particular, the receiver picks up two copies of
the transmitted signal,

$$s\left(t\right) \;=\; s_{\mathrm{TX}}\left(t\right) + 0.8 s_{\mathrm{TX}}\left(t - 2.5 T_s\right), \tag{186}$$

where the second copy is arbitrarily delayed from the first by 2.5 symbols. For
simplicity, all other parameters regarding the signal were kept the same. The effect
this distortion has on the PSD of the signal can be seen in Fig. 24. This figure shows
that the multipath interference both constructively, and destructively, interferes with
the signal of interest. In the worst case, that of destructive interference, this figure
shows that sections of the signal are almost completely wiped out.

The good news is that all of the formulas developed in Sec. 3.2 apply even in
severe multipath conditions. The only thing that needs to be discussed is $\psi\left(t\right)$. This

multipath environment is equivalent to receiving a signal that had been constructed with the pulse shape,

$$\psi(t) = \psi_{\text{TX}}(t) + 0.8\psi_{\text{TX}}(t - 2.5T_s). \tag{187}$$

The problem is that the receiver may not know what $\psi(t)$ is. What filter should be used in this case? If $\psi(t)$ is unknown, there are two options. Either it can be estimated or a filter can be generated based upon some assumed value of $\psi(t)$. Each method incurs a cost, the first in computational complexity and the second in performance. What the following experiment demonstrates is the ability to predict the performance loss associated with a non–optimal filter choice.

Four filters were tested in order to show this loss. The first two filters, called *mismatched* filters here, were formed under the (incorrect) assumption that the pulse shape had not changed since transmission. That is, they were designed under the assumption that there was no multipath interference. These filters are the matched filter for colored noise and the BPSK MMSE filter both shown in Fig. 15 on page 116. A second pair of filters, called *true* filters here, were created from the distorted pulse shape assuming perfect knowledge of the channel. The magnitudes of the transfer functions for these filters are shown in Fig. 25. The performance difference between these two pairs should highlight the importance of knowing the multipath channel.

To show this difference, we first examine the MSE at the output of the demodulator for each of the four filters. Given that the prediction methods are designed to predict MSE, this prediction should be exact. Comparing simulated results to predictions in Fig. 26, we see that the points, arrived at via simulation, do indeed match the lines showing the performance prediction. A second conclusion from Fig. 26 is that compensating for the multipath distortion is required in order to achieve a low MSE. This compensation needs to include not only the matched filter and over-whitener combination found in the True MF, but also the equalizer found in the
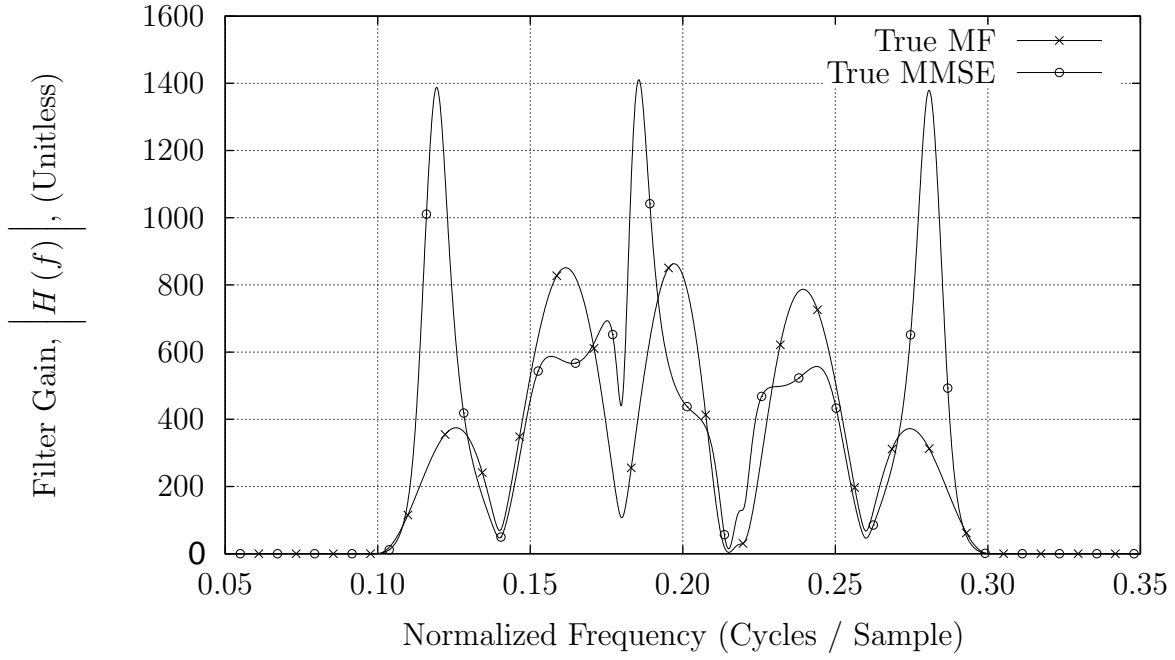
Figure 25.   Filters appropriate in a multipath environment
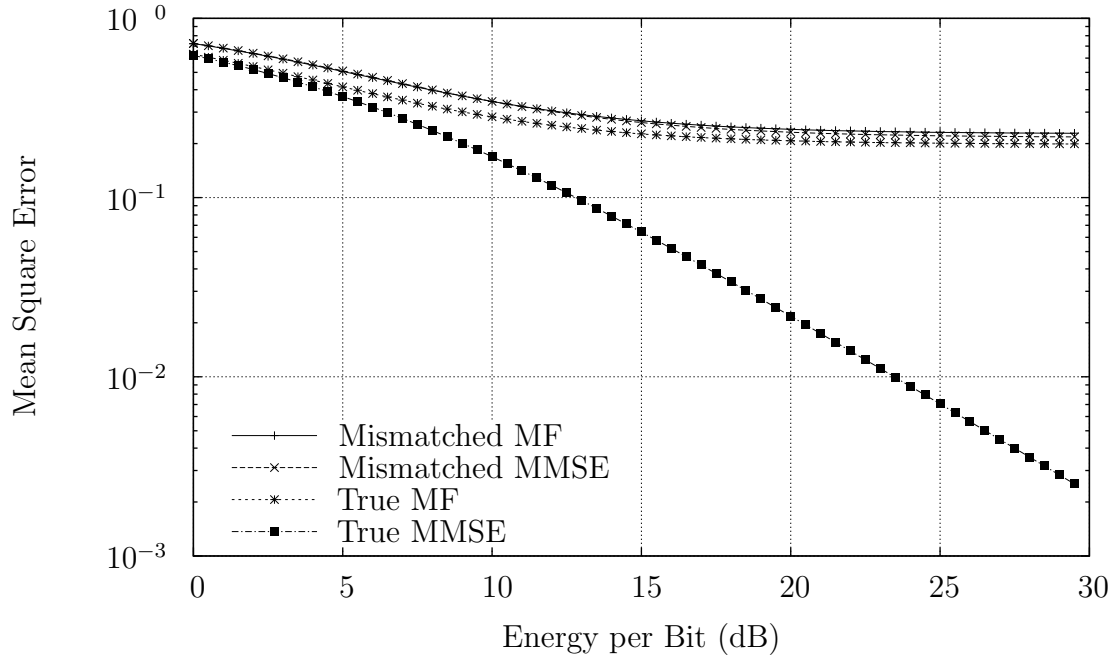
Figure 26.   Predicted Mean Square Error in Multipath Interference. The lines are the predictions, the points are arrived at via simulation.

128
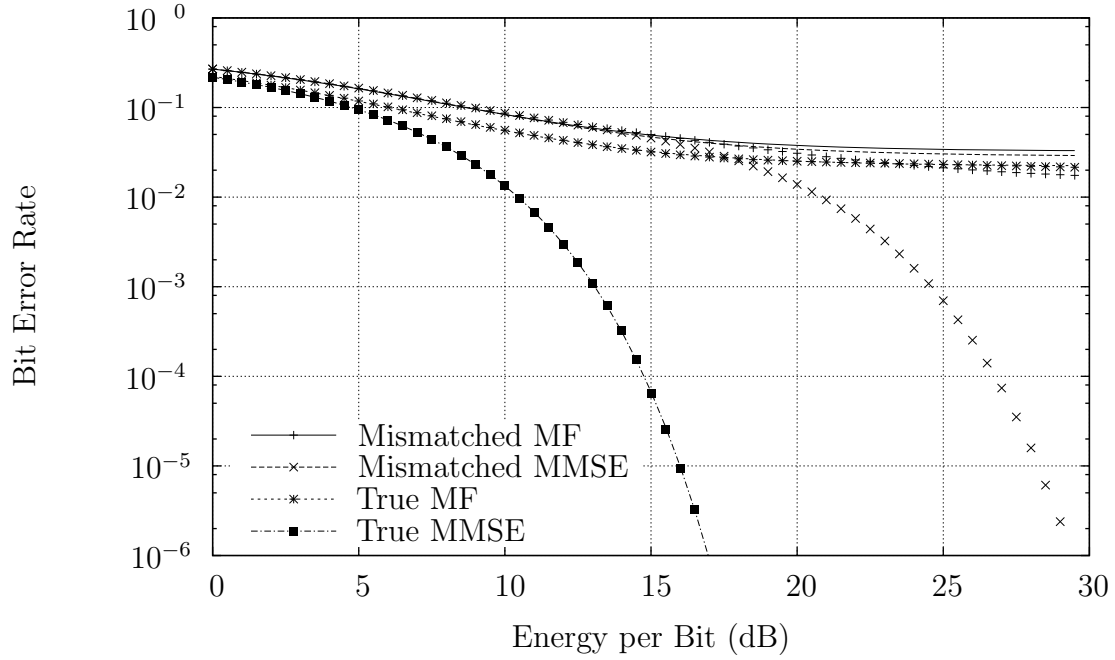
Figure 27.   Predicted Bit Error Rates in Multipath Interference. The lines are the predictions, the points are arrived at via simulation.

True MMSE. Without this equalizer, increases in signal strength do not necessarily decrease the error.

A more appropriate metric in system design, however, is the BER. Thus, even though the BER prediction formula given in Sec. 3.2.2 is only valid when the demodulator error is Gaussian, the BER was also calculated at the output of each demodulator and compared against the predicted BER. This result is shown in Fig. 27 where, as before, the points are the result of simulation and the lines are the result of the prediction formulas. Unlike the MSE predictions, however, the BER predictions are no longer accurate in every case. The one case where they are accurate, that of the True MMSE filter, corresponds to the one case where the disturbance is primarily Gaussian for small and large signal strengths. In the other cases, it appears as though the performance prediction is accurate for weaker signals and only departs as the signal gets strong.

One unusual feature of this graph is that the True MMSE filter in multipath distortion outperforms the BPSK MMSE filter when their is no multipath distortion (compare Fig. 27 with Fig. 23). This can be explained by the fact that multipath distortion results in more signal power reaching the receiver, from other directions, than could ever reach the receiver from one direction alone. Thus, if it could be compensated for properly, multipath distortion could be constructive.

When a system needs to operate in this type of environment, a system designer can use this performance prediction method to determine how much needs to be paid to achieve better performance. For example, if the system needed to a maximum 10% BER, it might be possible to achieve this by simply increasing the power. If, on the other hand, a BER less than 10^{-4} was required, something would need to be done to estimate the channel.

Either way the designer looks at the problem, these performance prediction methods provide one more tool for answering this question. The methods predicted the MSE performance exceptionally well. In terms of BER, the method only worked as long as the MSE was the result of a Gaussian disturbance. Even if the disturbance were not Gaussian, however, these methods still form a valid first order approximation into the BER that might be achieved. Better still, the performance of the true MMSE forms a lower bound on the performance achieved by any suboptimal demodulator.

4.2.3 Multi–sensor BPSK Reception.   Having demonstrated the capability of minimizing the MSE in a single–sensor demodulator, together with the capability of predicting the MSE from any single–sensor demodulator, we now turn to multi–sensor reception to continue the development of minimum MSE filters. The reason for a multi–sensor approach, or in this case a two–sensor approach, is simply to get rid of that persistent co–channel interference that plagues any system in a high interference environment. What a multi–sensor approach offers is the ability to

filter the incoming signals spatially in order to maximize the signal contribution and minimize the interference.

Yet common delay and sum beamforming methods are limited to removing no more than $M - 1$ interferers, or one less than the total number of sensors. In a high interference environment, however, the number of interferers may easily be larger than the total number of sensors, and these interferers may arrive from any direction. Therefore, it would be nice to have some sort of "optimum" filter, or even graceful degradation, as the number of interferers increases beyond this bound. As this section will demonstrate, the multi–sensor BPSK MMSE filter provides exactly that for BPSK signals.

Demonstrating this capability, however, requires an environment with at least as many interferers as sensors—otherwise no advantage could be shown over conventional delay and sum beamforming. If we confine ourselves to the two sensor filter presented in Sec. 3.2.4, then a minimum of two interferers are required to demonstrate this capability. To construct this interference environment, we start with the already familiar environment containing white noise and a narrow interference spike used in the single sensor filter tests of the last two subsections. To these two interferers, we add a broadband interferer covering the entire left half of the signal spectrum, creating the interference PSD shown in Fig. 28. The size, strength, and shape of these two interferers were chosen arbitrarily, save that they were designed not to overlap spectrally.

Further, in order to measure the gain associated with a two–sensor filter over a single sensor filter, the signal was placed on one of the two sensors only. That sensor, here referred to as the "horizontal" sensor to highlight the utility of this method on a cross–polarized antenna with vertical and horizontal feeds, shall receive both the signal, the noise, and the two interferers. This results in a received PSD equal to the one shown in Fig. 29. By placing the signal on one sensor only, the operation of single sensor demodulation has been simplified since any single sensor

0.16
0.18

0.22
0.24

0.26
0.28

0.35
0.40
0.45
0.50
−0.15
180
−0.10
160
−0.05
140
120

0.12
0.14

0.16
0.18
−20

0.22

0.24

0.26
0.28

0.35
0.40
0.45
0.50
180
160
140

−20

Power / Normalized Frequency

Normalized Frequency (Cycles / Sample)
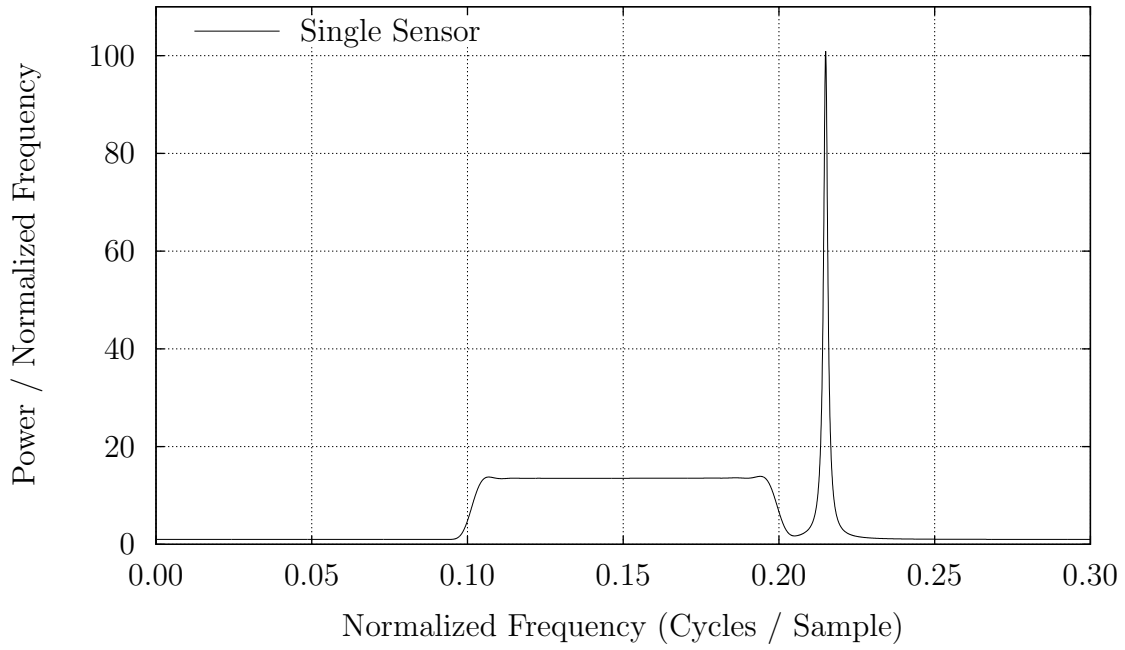
Single Sensor

Figure 28.   Single Sensor Noise PSD

Power / Normalized Frequency

Normalized Frequency (Cycles / Sample)

Signal PSD, $E_b = 15$ dB
Single Sensor Noise PSD
Received PSD
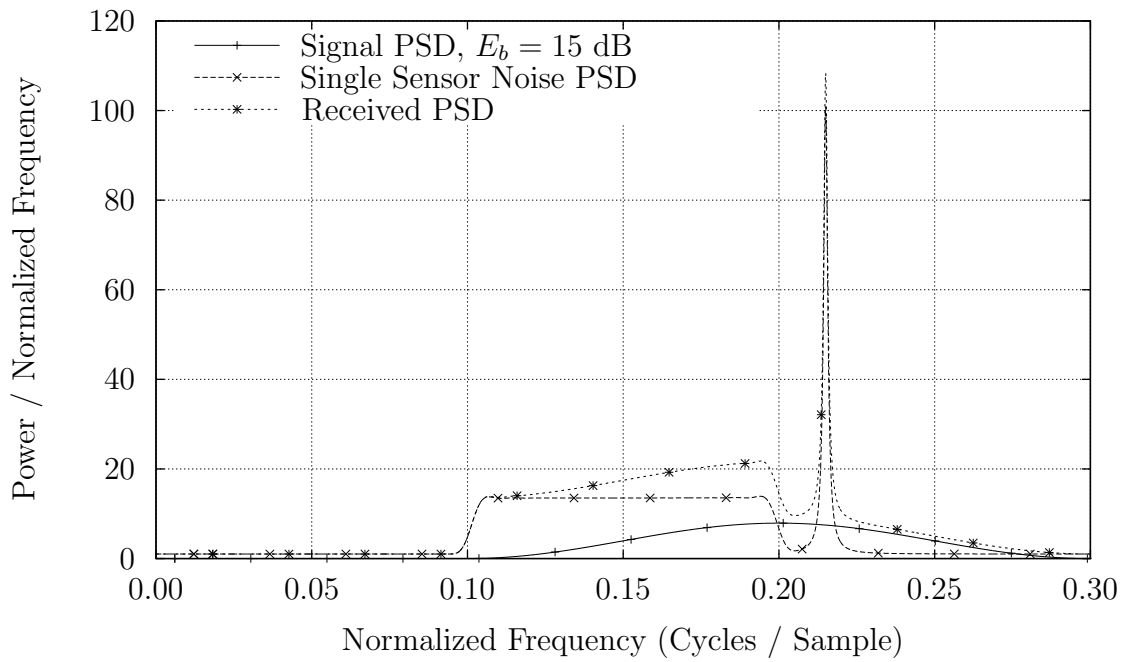
Figure 29.   Single Sensor Received PSD

132

demodulation method can spatially tune into the signal by simply discarding the second sensor. This can be thought of as an ideal beam former, tuning the antenna to the signal alone. As a result, comparisons between single sensor filters, corresponding to standard beamforming followed by filtering, and dual sensor filters are relatively straightforward.

Before proceeding any further, it should be pointed out that the two sensor filter derived in Sec. 3.2.4 is not limited to those cases where the signal is present on one sensor only while noise and interference are present on others. This latter limitation, common to most adaptive algorithms [70], is not appropriate here. Placing the signal on only one sensor simply makes it easier to test and compare the dual sensor filter with the equivalent single sensor filters.

Having now described an environment sufficient to design and implement single sensor filters, we turn our attention to examining these filters. In particular, the four filters tested earlier in Sec. 4.2.1 will form four filters for comparison in a dual sensor environment. These filters are shown in Fig. 30, and their operation is much the same as before even though their appearance has been altered to compensate for the additional broadband interferer.

From here we proceed to describe the environment received on the second sensor. Careful examination of the two sensor filters in Eqns. (114) and (115), shows that without any correlation between the interferers received on each sensor the two sensor filter is equivalent to a single sensor filter. Therefore the second sensor, or the "vertical" sensor as it is called here, will receive both of these interferers as well. Then, in order to keep this demonstration different from common spatial developments, these interferers will come from different spatial directions—creating a negative correlation between the sensors for the broadband interferer, and a positive correlation for the narrowband interferer. This creates the cross spectral density, as defined by Eqn. (5) on page 13, shown in Fig. 31, while maintaining the previous single sensor PSD shown in Fig. 28. Because this correlation is not constant across the
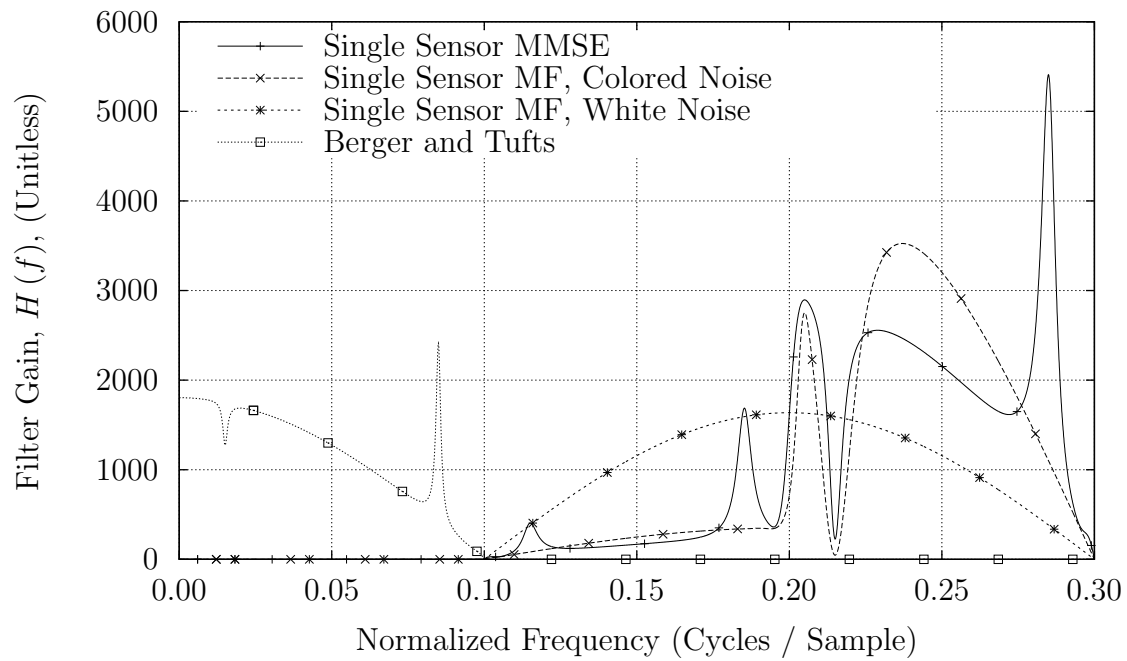
Figure 30.   Single Sensor Filters used in the Two–Sensor Test
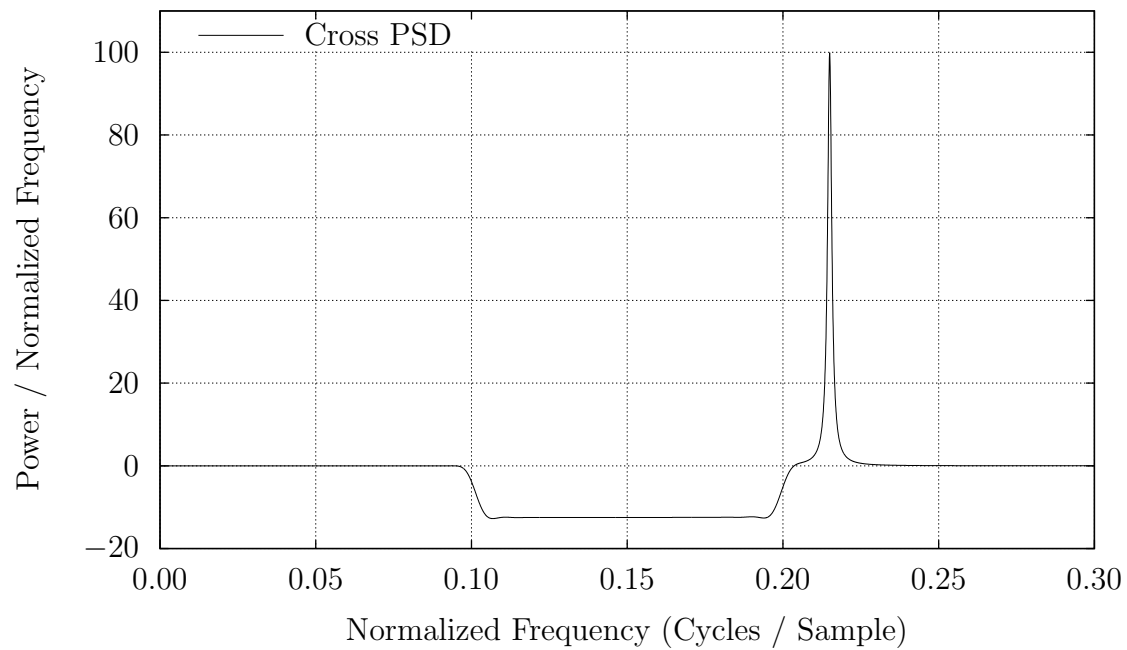
Figure 31.   Cross Power Spectral Density

134
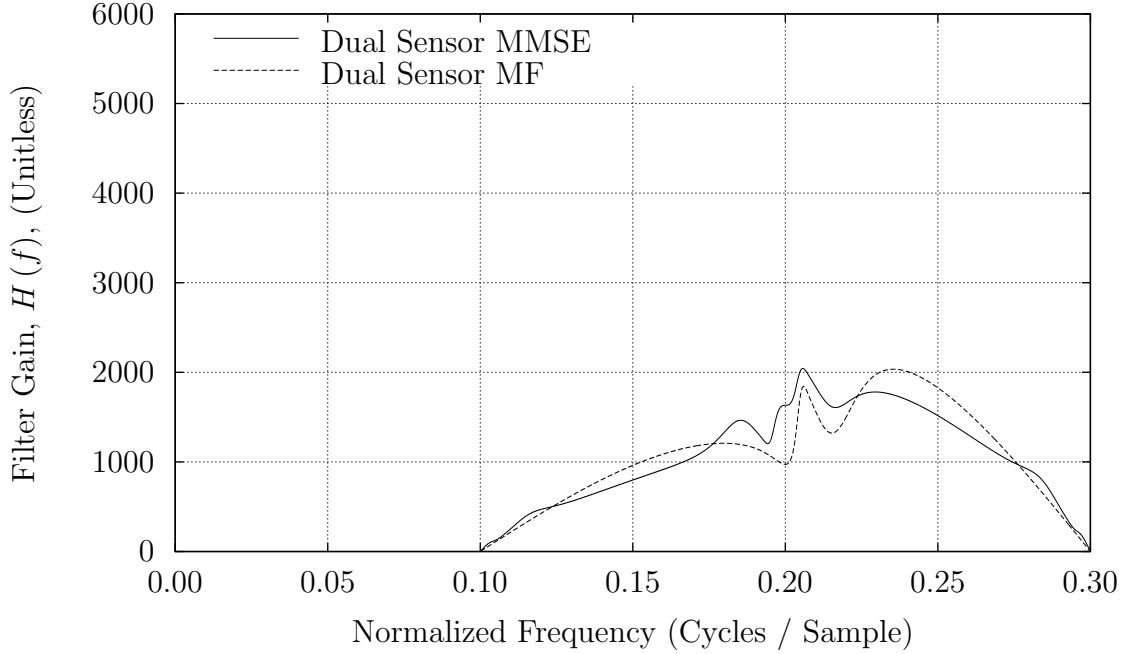
Figure 32. Horizontal Two–sensor Filter Component

band, it is impossible to spatially notch one direction and filter out both interferers. Indeed, any attempt at spatially notching a single direction will result in increasing the interference from the other direction.

Having developed a two sensor signal and interference environment, the next step is to introduce the two sensor filters applied in this environment. Two dual sensor filters were tested here. The first filter is the two–sensor BPSK MMSE filter derived in Sec. 3.2.4 and found in Eqns. (114) and (115). The second filter is the two sensor matched filter, defined similar to the BPSK MMSE filter save that $H_{\text{EQ}}(f)$ in Eqn. (116) is arbitrarily set to 1. This filter is derived in [35] and commonly used in Radar applications. These two filters are shown in Fig. 32, corresponding to the filter applied to the "horizontal" sensor containing the signal, and Fig. 33, corresponding to the filter applied to the sensor that is not receiving the signal. The first, and most obvious difference, between these filters and the single sensor filters is that a filter is applied to the vertical sensor even though no signal is present on

135
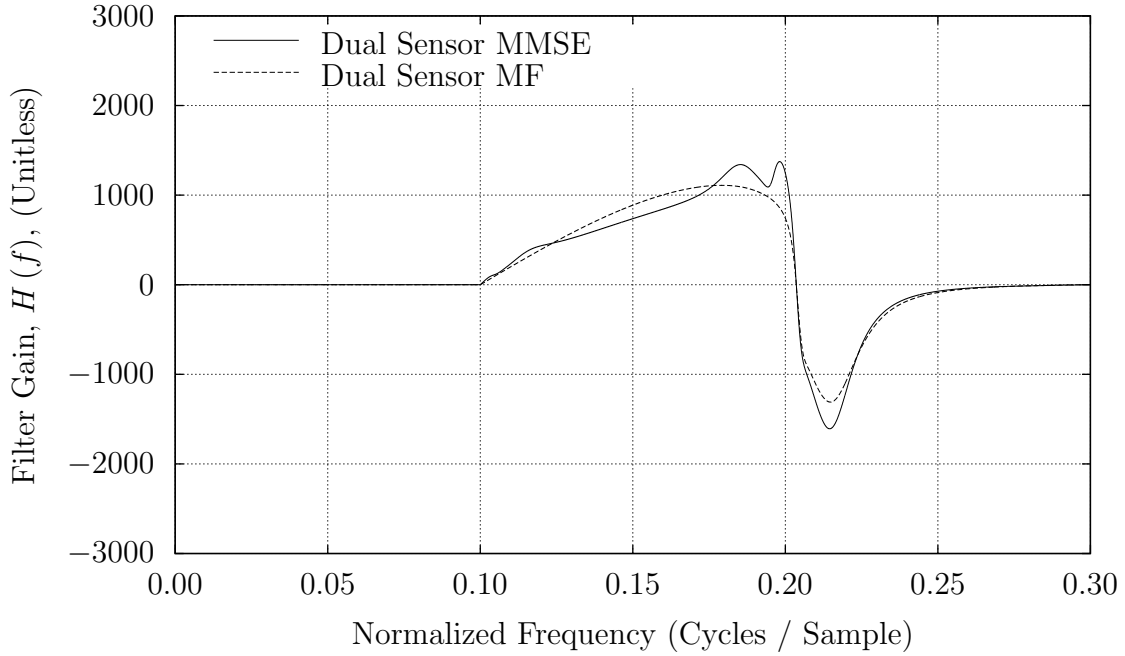
Figure 33. Vertical Two–sensor Filter Component

this sensor. This extra filter helps to remove the interference from the system at the front end. Indeed, so much interference is removed at the front end that the equalized filter, that is the two sensor MMSE, doesn't appear significantly different from the two sensor MF.

To understand the operation of these two multi–sensor filters, we must compare them with the cross spectral density of the interference given in Fig. 31. Note that when the broadband interferer is present, between 0.1 and 0.2 CPS, the filter response is positive on both sensors. This effectively notches out the wideband interferer spatially across that part of the spectra. Examining the narrowband interferer, at 0.215 CPS, these two filters subtract the signals as they are received from each sensor. Like the previous response for the broadband interferer, this subtraction combines to spatially notch out the interference. Further, as the frequency increases and the strength of the narrowband interferer diminishes, the response on the second sensor

Figure 34.   Dual Sensor Mean Square Error. The lines for the two MMSE filters are predictions, and the points are arrived at via simulation. All other lines are arrived at via simulation.

goes to zero, effectively maximizing the signal strength where no interferer is present. This is something that common beamforming alone cannot do.

To show the performance of this two sensor filter in this interference environment, we start by examining the MSE at the output of the demodulator following each filter, as shown in Fig. 34. Starting with the single sensor filters, their ordering and performance is similar to the performance seen in the single sensor filter test in Sec. 4.2.1, although some differences are worth mentioning. The first difference is that Berger and Tufts' filter performs little better than the matched filter for white noise. This is a result of the fact that the broadband interferer can only be removed prior to downconversion. After downconversion, the broadband interferer has corrupted the entire band of the signal and Berger and Tufts' filter is helpless to compensate. A second difference is seen in the performance of the colored noise matched filter. The ISI suffered by this filter is about four times stronger than be-

fore, preventing this filter from achieving good performance as the signal strength increases. As might be predicted, the single sensor BPSK MMSE filter continues to outperform all other single sensor filters.

Further observations may be made regarding the two–sensor filters. The first, most obvious, observation is the these two filters see less MSE than their single sensor counterparts. Measuring this difference shows that it is about 3.4 dB less (see arrow in Fig. 34). Second, the performance difference between the dual sensor matched filter and the dual sensor BPSK MMSE filter looks remarkably similar to the difference between their single sensor counterparts. The dual sensor matched filter clearly removes most of the noise, yet suffers from a similar limiting problem with ISI, while the dual sensor MMSE filter apparently has no such problem with ISI. This commends the dual sensor BPSK MMSE filter over and above the dual sensor matched filter.

The final performance measure between these two filters is the BER performance shown in Fig. 35. As with the MSE, the two sensor MMSE filter outperforms all others in terms of BER as well. Likewise the BER performance of Berger and Tufts' filter follows its MSE performance. The matched filters are interesting here, however, primarily because the single sensor matched filter does much worse than it did in the single sensor experiment in Fig. 19. Knowing this suggests that, in other environments, the performance of the two sensor matched filter may likewise be much worse than this experiment demonstrates just as the performance of the single sensor matched filter in Fig. 19 became much worse here as the environment changed.

The performance seen in Figs. 34 and 35 needs to be caveatted by the fact that the gain a MMSE filter achieves over another filter is interference dependent. For example, in a purely white noise environment with no correlation across the sensors, the BPSK MMSE filter and Berger and Tufts' filter both converge to a matched filter, providing no gain. This is why the performance prediction equations for the single
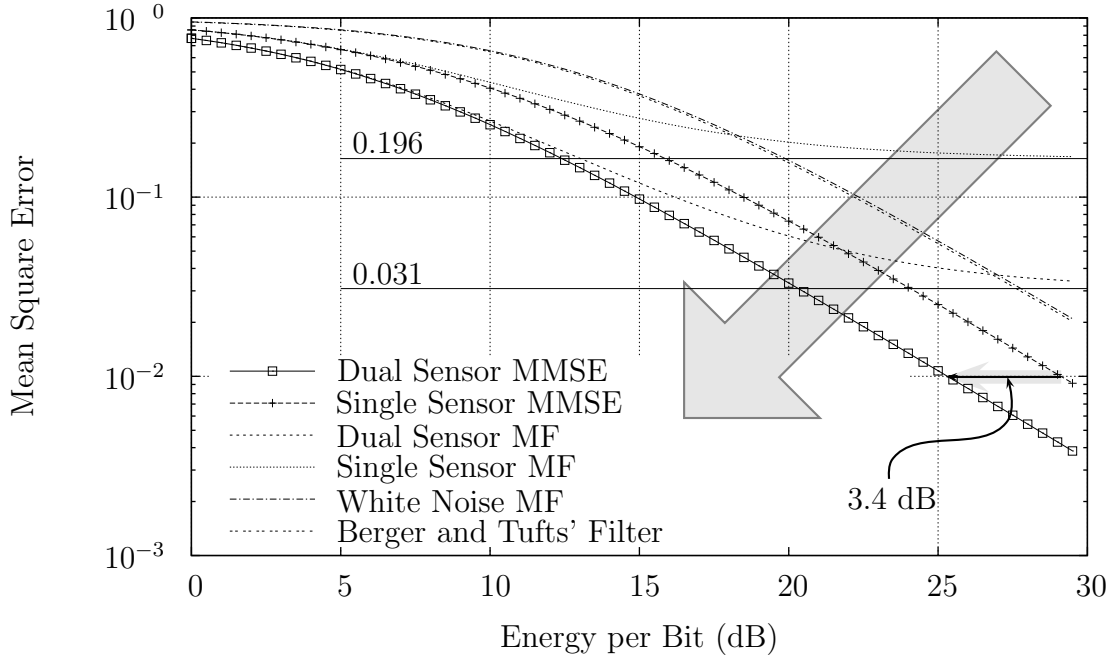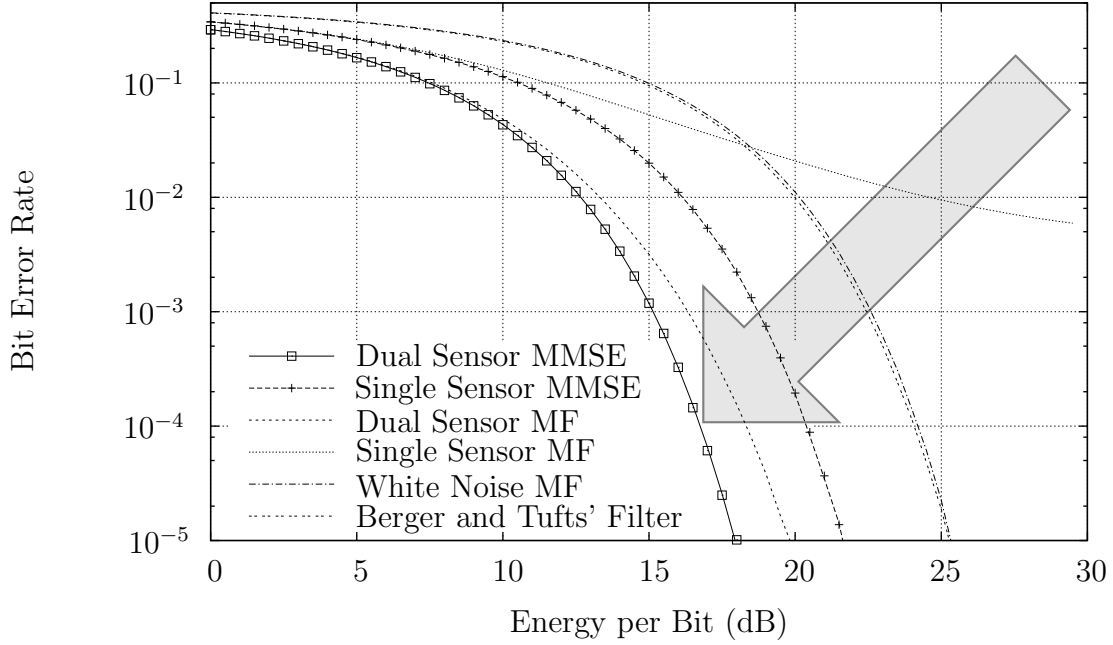
Figure 35.   Dual Sensor Bit Error Rate. The lines for the two MMSE filters are predictions, and the points are arrived at via simulation. All other lines are measured.

BPSK MMSE filter in Eqn. (81) and for the dual BPSK MMSE filter in Eqn. (118) are so important. These equations make it possible to predict the performance of these filters in any environment. Just to show how well these equations work, the lines corresponding to the MMSE filter performance shown in Figs. 34 and 35 correspond to the results of using the prediction formulas, while the points are the results of simulation. The two are exactly on top of each other, validating that these formulas are indeed accurate. That means that, given an interference environment, these formulas may be applied to calculate the performance one might expect from a MMSE filter without resorting to simulation.

Finally, the performance of this dual sensor MMSE filter clearly shows the advantages of MMSE filtering over delay and sum beamforming. By adding the second sensor, this simulation demonstrates that correlated noise between the sensors can be used to remove the interference from the channel. Further, this experiment helps to highlight the use of an equalizer again. In this two sensor case, the performance

139

gain from the equalizer alone came close to 2 dB for the two sensor filter at a BER of 10^{-5}, while similar BER's were unattainable from the single sensor matched filter.

Looking back upon every BPSK MMSE test applied in this section, they have all demonstrated that the most important place to filter out the noise is prior to combining the redundant components. That is, the noise needs to be mitigated prior to combining sensors in the case of a two sensor reception problem, prior to downconversion, and prior to sampling the output signal. Only by filtering the noise properly at the front end is good performance obtained.

4.3   Time Difference of Arrival Estimation

Having now derived and demonstrated a valid estimate of the unknown data, the next step is to demonstrate the utility of an estimator based upon this data estimate: the TDOA estimator. While filtering is primarily concerned with determining *what* was said, TDOA estimation is a first step in determining *where* it was said. In particular, the estimator derived in Sec. 3.3 will be demonstrated here in two environments. The first is the standard white noise test environment where this new TDOA estimator will be shown to outperform all of the estimators presented in Chapt. II. Then, following the white noise test will be two colored noise tests, similar to Streight's tests in [61]. These tests are especially interesting since cyclostationary TDOA estimates supposedly achieve their best performance in strong interference environments. This section will show that the ML TDOA estimate derived in Sec. 3.3 would outperform Streight's ideal MLE if it could be implemented. The final subsection will then present the loss associated with the suboptimal lopsided TDOA estimator in white noise, as a function of how well its assumptions were met.

Especially important, within this section, is that this is the first cyclic TDOA estimation work to openly compare optimal stationary TDOA estimators with optimal cyclostationary TDOA estimation algorithms. While previous cyclostationary TDOA estimators have been consistently compared against the GCC, they have not

been compared against the Eckart filter. For example, Gardner and Chen [24], as well as Streight [61], compare their estimators against the GCC only. Neither of these previous works, or even others following such as [14], have compared the performance of their algorithms to either the stationary MLE or the Eckart filter. Not having a proper comparison between these methods leaves their relative performance unknown. The results presented in this section highlight just how important such relative performance comparisons are.

4.3.1 Implementing the ML TDOA Estimator. The first step to illustrating the performance of these estimators is to test them in a white noise environment. While this is certainly a standard test environment, this test marks the first test of a practical low–SNR cyclic MLE.[7] In particular, two tests will be conducted in this white noise environment. The first test will demonstrate the performance of the cyclic TDOA estimator as a function of SNR, and the second will demonstrate the same performance as a function of the observation length. In addition, this second test will be used to explain performance differences between those reported in [24] and those presented here. These differences will be traced to a failure to do subsample interpolation in previous studies.

The first order of business, however, is to present the experimental approach and then the signal parameters. For this experiment, one signal and two white noise sequences will be generated. The signal will be added to one noise sequence to create the first sensor's input, as well as delayed by τ_d and then added to the other noise sequence to create the second sensor's input, as shown in Fig. 36. The result of this process will then be grouped into blocks of length $T = N_s T_s$, and an FFT will be applied to each block. This block FFT output will then be placed into the TDOA estimator under test. Each TDOA estimator will then produce a function,

---

[7]As noted in Sec. 2.3, Streight's low–SNR cyclic MLE was only "ideal," and therefore could not be implemented. The low–SNR cyclic MLE presented in Sec. 3.3, by contrast, can be implemented since the spectral correlation function is assumed to be known to within a complex scale constant.
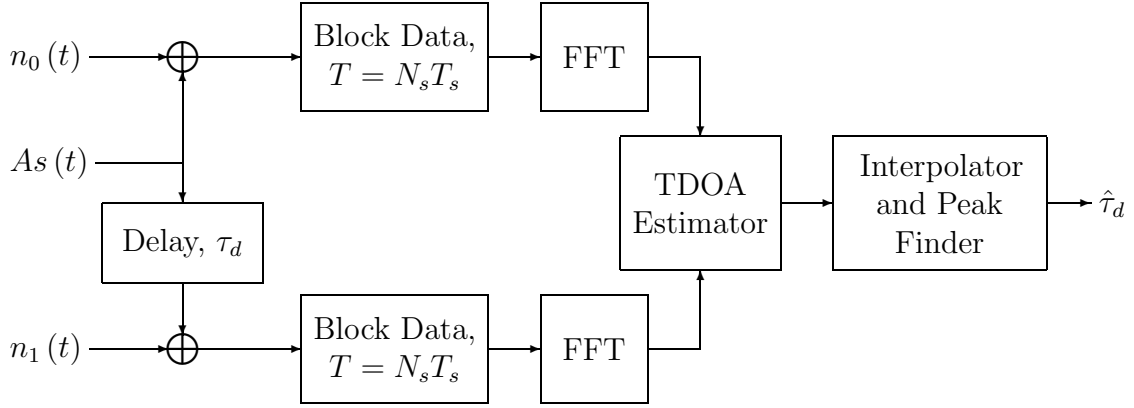
Figure 36. TDOA Test Setup

such as the $g[\tau_d]$ function described in Eqn. (139) for the cyclic MLE, which needs to be maximized. The location of the maximum of this function then becomes the estimate of the TDOA. Finally, the performance of each routine will be measured using the mean square error, which will be calculated as,

$$
\text{MSE}_{\text{TDOA}} \text{ dB} \quad \approx \quad 10 \log_{10} \left[\frac{1}{T_s N} \sum_{n=0}^{N-1} (\hat{\tau}_{d,n} - \tau_d)^2 \right], \tag{188}
$$

where N is the number of test cases examined and $\hat{\tau}_{d,n}$ is the TDOA estimate for the nth test.

The last step before introducing test results is to define the signal environment used for these tests. Starting with the signal, the signal of interest will be a burst QPSK signal since the estimator derived in Sec. 3.3 was specific to the redundancies found within a QPSK signal. Second, given that the cosine pulse function defined in App. A is easy to generate and has low sidelobes, this pulse function will be used to generate the signal. The carrier of this signal has been placed at 0 CPS for convenience, and its symbol rate has been set to 0.125 CPS corresponding to a symbol length of 8 samples. These parameters place the signal squarely within the complex frequency band, while keeping the effects of sampling to a minimum. One unintended consequence of this choice, however, is that all of the noise outside of a

142

Table 4.   Signal Parameters for TDOA Estimation Tests

Signal Parameter	Parameter Value
Modulation Type	QPSK
Carrier Frequency, f_c	0 CPS
Symbol Rate, $\frac{1}{T_s}$	0.125 CPS
Pulse Function, $\psi(t)$	Cosine
Relative Magnitude, $\frac{A_0}{A_1}$	1
True TDOA, τ_{d1}	3 T_s
Phase Mismatch, θ_{d1}	0
Noise Covariance, $S_n(f)$	Constant (white)

band between -0.125 and 0.125 CPS, will contribute to the poor quality of the GCC estimate—which does nothing to remove this out of band noise. Finally, the phase mismatch will be arbitrarily held to zero and the magnitudes on both sensors will be arbitrarily kept identical. This leaves the last parameter, the true TDOA value, which was set to $3T_s$ to match the test cases given in [24]. These parameters are summarized in Table 4.

Having covered the background, the first test measured TDOA estimation error as a function of SNR. The results from this test, in terms of MSE vs. SNR, are shown in Fig. 37, corresponding to an observation length of 128 symbols, and Fig. 38 corresponding to an observation length of 1024 symbols.[8] The MSE shown is measured in dB's with a 0 dB MSE corresponding to a MSE of one symbol, while -24 dB corresponds to a single sample error. Table 5 summarizes the relative capability of the algorithms shown in Figs. 37 and 38.

The first feature apparent from these figures is that there is gentle curve followed by a sharp bend in each TDOA estimation plot. These features are common

---

[8]The results in Figs. 37 and 38 follow averages of 2,723,840 and 3,899,392 trials respectively. This excessive number of trials was necessary to remove the effects of the high variance surrounding the bend in the plot.
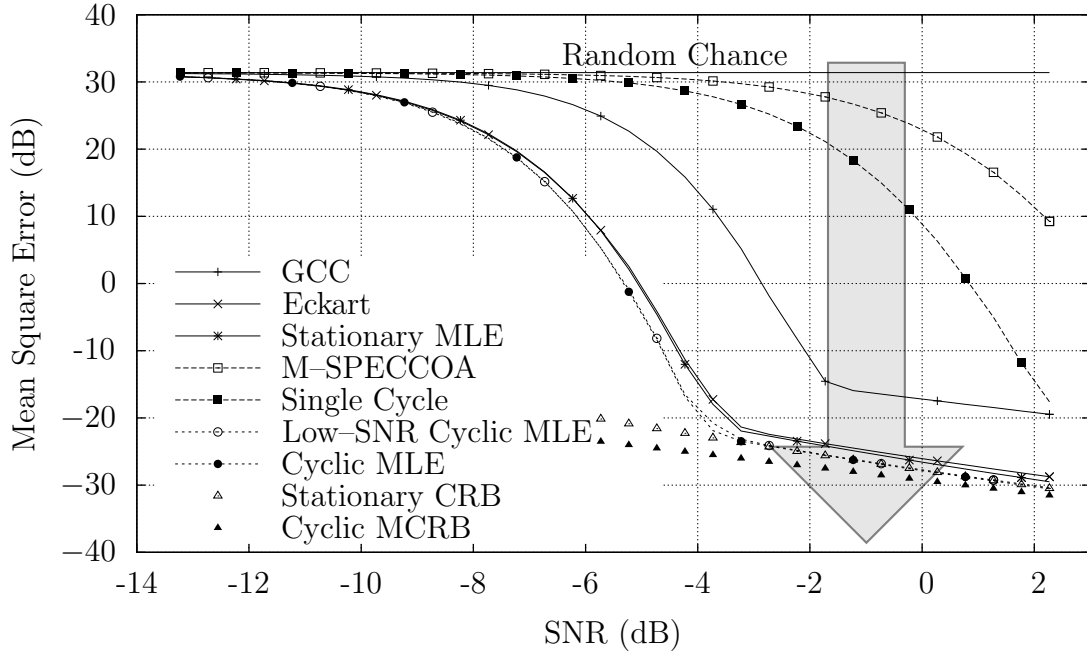
Figure 37. TDOA Estimation Error for a 128 Symbol QPSK Burst
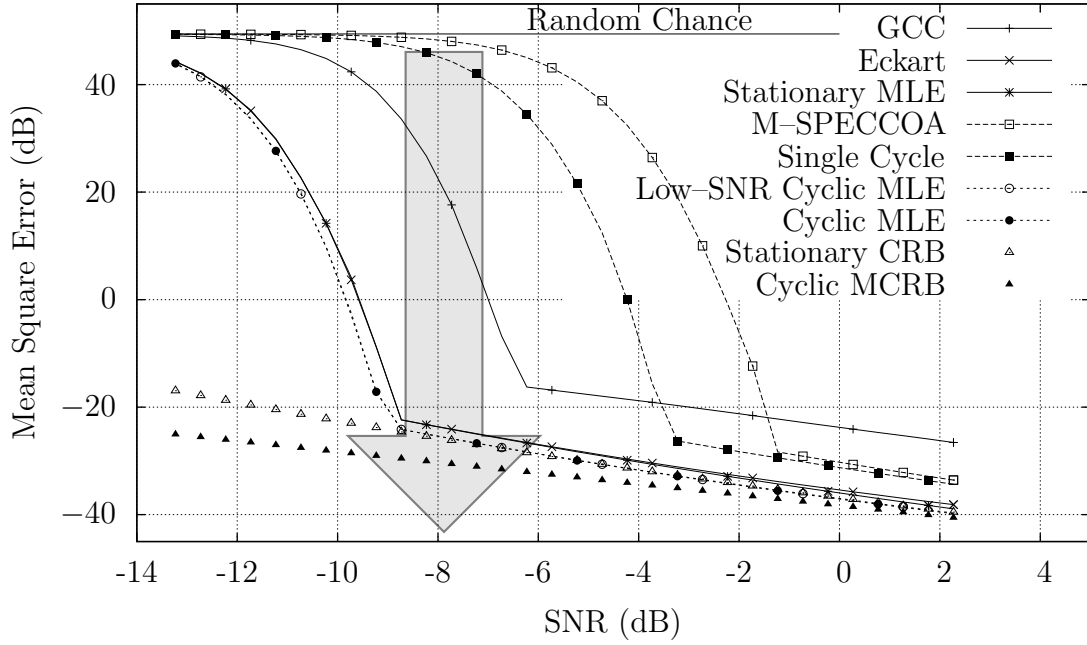
Figure 38. TDOA Estimation Error for a 1024 Symbol QPSK Burst

144

Table 5.   TDOA Estimator Ranking

Rank	Name	Equation	Page
1.	Cyclic MLE	126	83
2.	Low–SNR Cyclic MLE	139	88
3.	Stationary MLE	24	26
4.	Eckart	23	26
5.	GCC	21	25
6.	Single cycle	147	38
7.	M–SPECCOA	144	92

in maximum likelihood estimates where a search is required over the parameter of interest, and they have been documented and studied elsewhere [69, Sec. 10.2.2]. The curve can be explained by looking at the effect of low SNR on TDOA estimation. For exceptionally low SNRs, the signal provides no TDOA information at all and so the best estimator is no better than a uniform random number generator. Working quickly from first principles, such a random number generator would have a mean square error given by,

$$\mathcal{E}\left\{(\hat{\tau}_d - \tau_d)^2\right\} = \int_{-\frac{T}{2}}^{\frac{T}{2}} (\hat{\tau}_d - \tau_d)^2 \frac{1}{T} d\hat{\tau}_d = \frac{T^2}{12} + \tau_d^2. \tag{189}$$

This random chance MSE is annotated on each plot as a line near the top. As the SNR increases, the probability of a random chance distribution lowers as it becomes more and more likely that the estimated value is close to the true one, causing the downward curve in the performance charts. When the probably of a totally random result reaches near zero and the probability of being close to the true answer nears one, the probability distribution governing the TDOA estimate becomes dominated by the local variation instead of spurious peaks located throughout the TDOA search range. At the point where this switch takes place, the probability distribution of $\hat{\tau}_d$ changes fundamentally, causing the sharp bend in each curve. Following this final change the performance appears to become linear.

Second, the Cramér–Rao bound for stationary signals seems reasonable to the right of the sharp bend, but for low SNRs that same bound seriously underestimates the error. This is also a known phenomenon discussed in [69]. Knapp and Carter explain this phenomenon with respect to TDOA estimation by saying that,

> It should be pointed out that [the variance of an arbitrary estimator] and [the Cramér–Rao bound] evaluate the *local variation* of the time–delay estimate and thus do not account for *ambiguous peaks* which may arise when the averaging time is not large enough for the given signal and noise characteristics. Indeed, when T is not sufficiently large, local variation may be a poor indicator of system performance and the envelope of the ambiguous peaks must be considered. [37, p. 325] [Emphasis added]

Likewise the Cramér–Rao bound formulas given in Sec. 3.3.3 are only estimates of the "local variation" in their respective estimators. When the SNR is too low or the observation length too short, these local estimates provide no indication of the true estimation capability. This explains why the performance of the Eckart filter and cyclic MLE methods track closely with the CRB for high SNR, only to leave that bound as the SNR gets much lower.

Another comparison may be made between the Low SNR Cyclic MLE and the true Cyclic MLE, and likewise the difference between the Eckart filter and the stationary MLE. If you will recall from Sec. 2.3 and likewise Sec. 3.3, the true MLE estimate requires knowing the strength of the signal. Since this is not necessarily known at the receiver, any implementation of a true maximum likelihood estimator must either attempt to estimate the signal strength or use one of these approximations. Looking at Figs. 37 and 38, the difference between the performance of the low–SNR approximations, as compared to the true estimators, is small. Of these differences, the greatest difference is between the stationary MLE and the Eckart filter. The cyclic TDOA MLE, on the other hand, does not appear visibly better than the low–SNR approximation of it on this chart at all. Judging from these small differences, there appears little reason to expend the resources necessary to estimate the true amplitude of the signals in question—especially when cyclic TDOA estima-

146

tion is desired. While estimating the true signal strength may provide a small gain in TDOA estimation, the gain demonstrated here is likely to be so small that it is insignificant. This justifies the low–SNR approximation for all SNRs.

A fourth feature worth noticing is the difference between the stationary and cyclostationary methods. As you will recall, comparisons between these methods, and in particular comparisons between Eckart's filter and cyclostationary TDOA estimators, have not previously been published. As their relative performance has previously been unknown, Figs. 37 and 38 demonstrate for the first time that, while the cyclic methods improve upon the Eckart filter, this improvement is minimal (15–20%). Given that the low SNR cyclic MLE contains a term identical to the Eckart filter term, the $B_0[\tau_d]$ term in Eqn. (130), it would appear as though the other terms in this MLE are not as critical to the TDOA estimate as the Eckart filter term. Given this information, it comes as no surprise that the two single cycle estimates, composed of the $\alpha = \frac{1}{T_s}$ terms only, do much worse than the Eckart filter. These comparisons have been missing in previous cyclostationary TDOA developments.

Finally, the performance of the GCC is decidedly worse than that of the single cycle methods for large SNRs. This can be directly attributed to the fact that the signal is limited to a bandwidth of $\frac{2}{T_s} = 0.25$ CPS, while the white noise sequence occupies the full 1 CPS band. Since the GCC method does not use any filtering, it accepts the whole band of noise. This is unlike the single cycle TDOA methods, which only examine that portion of the input frequencies containing the signal of interest. This allows the single cycle methods to outperform the GCC once they get to the strong side of the bend.

These results contradict two conclusions presented within [24] and [61]. First, single cycle TDOA estimators do not, in general, outperform stationary TDOA estimation algorithms—such as the Eckart filter which wasn't tested in [24] or [61]. Second, these same single cycle methods do not outperform the Cramér–Rao bound given in [37].
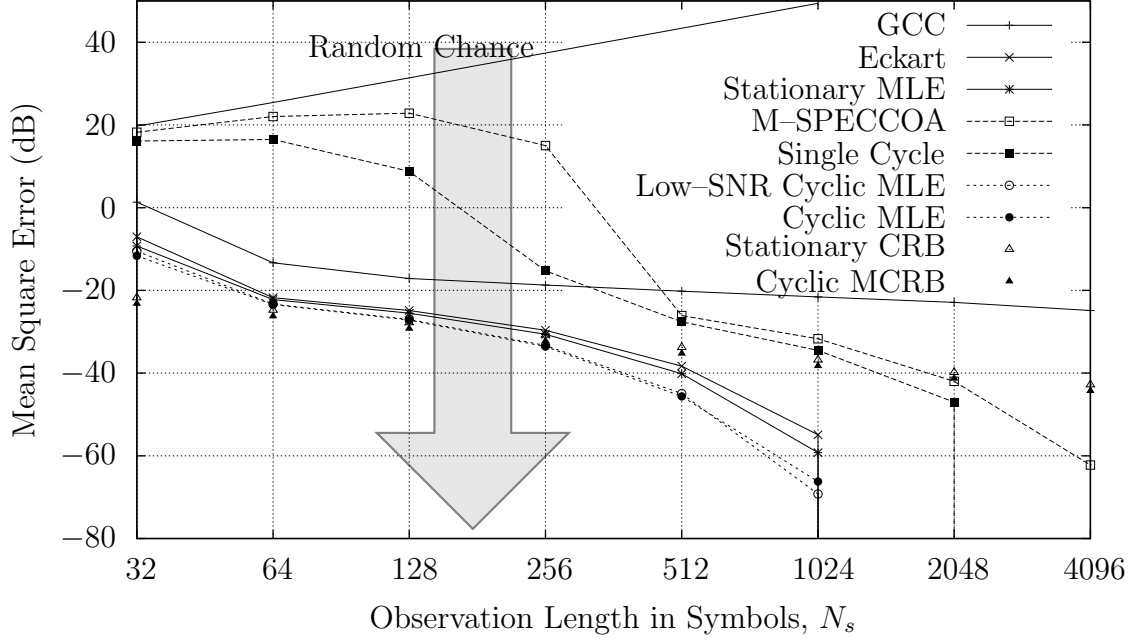
Figure 39. TDOA Comparison without Interpolation

The rest of this subsection will chase down the difference between these previous results and the results presented in Figs. 37 and 38. In particular, the next test, shown in Fig. 39, shows that similar misleading results can be achieved by repeating a similar experiment using a 0 SNR signal without applying any form of subsample interpolation. This test is somewhat modeled after the tests in [24] and [61], and thus it shows MSE performance as a function of the observation length.

Fig. 39 also supports the erroneous contention that cyclostationary TDOA estimators outperform the Cramér–Rao bound for stationary signals. This claim was been made under the assertion that, "the conventional CRLB does not apply ..." [24, p. 1193]. Because this assertion is quite valid, the new cyclic MCRB derived in Sec. 3.3.3 is presented as well here. Although the MCRB is not the true CRB, it is guaranteed to be lower than the true CRB. Therefore performance below this latter bound, such as the performance shown in Fig. 39, indicates a mathematical impossibility. Perhaps if previous results had included comparisons to the Eckart

filter this impossibility would have been more obvious, since the conventional CRB applies to the performance of the Eckart filter.

To Gardner and Chen's credit, they do point out that,

> Experimentation showed that 400 Monte Carlo trials yielded adequately stable values of MSE up to a point, that is, as long as the normalized MSE remained above about 40 dB. For normalized MSE below 40 dB, a larger number of trials would be necessary, and *time interpolation between sample values in the TDOA parameter would be advisable.* [24, p. 1194] [Emphasis added.]

While it was not possible to repeat Gardner and Chen's experiment exactly, and so the units in this chart do not necessarily match theirs, this quote still points out that they were uncertain of the exceptionally low MSE results they obtained.

This erroneous result can be explained by the fact that the true TDOA in this experiment, as in [24, 61], is an integer number of samples. Each estimate of this TDOA, without subsample interpolation, is an integer number of samples as well. When all of the estimates land on the same sampled value, the result indicates a zero measurement error. This leads to the measurements shown in Fig. 39, and the potentially erroneous conclusions which accompany this figure.

Fig. 40 is offered for comparison. The only difference between Fig. 39 and Fig. 40 is the use of the quadratic interpolation method given in Eqn. (140). From this new figure, we see that the single cycle estimators no longer outperform the Cramér–Rao bound. This demonstrates that both of these conclusions were the result of a failure to do subsample interpolation. Single cycle estimators do not outperform the CRB for stationary signals and they are not, in general, better than stationary methods.

As an added benefit, Fig. 40 illustrates the importance of the observation length. Improvements in observation time are roughly equivalent to improvements in SNR. This means that there are two ways of increasing the performance of a TDOA estimator. The first is to increase the SNR of the respective signals, while

the second is to increase the observation length. Barring either improvement, the best algorithmic increase possible is to use the cyclic TDOA MLE.

From all of these white noise experiments a few conclusions are in order. First, the cyclic MLE does truly outperform the stationary MLE. However, it only outperforms this estimator by about 15–20% in white noise. Second, the single cycle TDOA methods, which are supposedly completely immune to noise and interference, perform much worse than the Eckart filter—a comparison not made previously. This suggests that, while spectral correlations for non–zero cycle frequencies are truly signal specific, such spectral correlations are not the strongest features present in the signal. The strongest features are those in the zero cycle correlations. While cyclic TDOA estimation offers improvements beyond what is possible with a stationary TDOA estimator, such improvements in this case are only minimal.

4.3.2   TDOA Estimation in Interference.   Having examined the performance of several TDOA estimators in a white noise environment, and the necessity of using some form of subsample interpolation, we now turn to examining their performance in a high interference environment. Such comparisons are especially important because cyclostationary methods presumably outshine all others in these environments. In particular, the impact of the overwhitener, deemed unnecessary by other cyclic TDOA estimator developments, will be examined here.

This section will test the algorithms from the last section in two colored noise environments, each drawn from test cases presented by Streight [61]. The first environment contains wide band interference, while the second contains narrow band interference. Each of these two tests forces a non–unitary overwhitener, $\mathbf{R}_n^{-1}$. Although the form of this overwhitener will change from one environment to the next, both environments will highlight its impact. The results of the simulations presented here will clearly demonstrate that tremendous gains can be had in high interference environments by simply using an overwhitener.

Table 6.   Simulation Parameters for Colored Noise TDOA Testing

	Signal of Interest	Wideband Interferer	Narrowband Interferer
Carrier Frequency, f_c	$\dfrac{63}{256}$	$\dfrac{54}{256}$	$\dfrac{54}{256}$
Symbol Rate, $\frac{1}{T_s}$	$\dfrac{1}{16}$	$\dfrac{1}{10}$	$\dfrac{25}{1000}$

The two noise environments compared here are similar to those used by Streight [61]. The parameters he chose are shown in Table 6. Since the carrier frequencies Streight considers are non–zero, this will mark the only set of tests within this dissertation that focus on *bandpass* TDOA estimation. One key difference sets this testing environment apart from Streight's actual test cases, and that is that the Frequency Difference of Arrival, also known as the Differential Doppler, for all signals will be zero. That means that, in order to maintain the uncorrelated interference assumed in Sec. 3.3 while maintaining the same interfering signals on both sensors, the interference TDOA values need to be greater than the largest observation window.

That leaves the relative signal and interference strength parameters undefined. To set these parameters, we again follow Streight's lead, only we examine his worst case scenarios only—since these are the most interesting ones anyway. Specifically, Streight held the signal to white noise ratio constant at 0 dB, while the signal to interference plus noise ratio, SINR, was varied upwards from -9 dB. These two values, 0 dB SNR and -9 dB SINR will be used here. This places the TDOA estimator at a great disadvantage, having little signal strength to work with compared to the noise and interference.

All of these parameters together combine to create the noise plus interference and signal power spectral densities shown in Figs. 41 and 42. Because the environments are so poor, these PSD plots show the signal strength being severely dwarfed
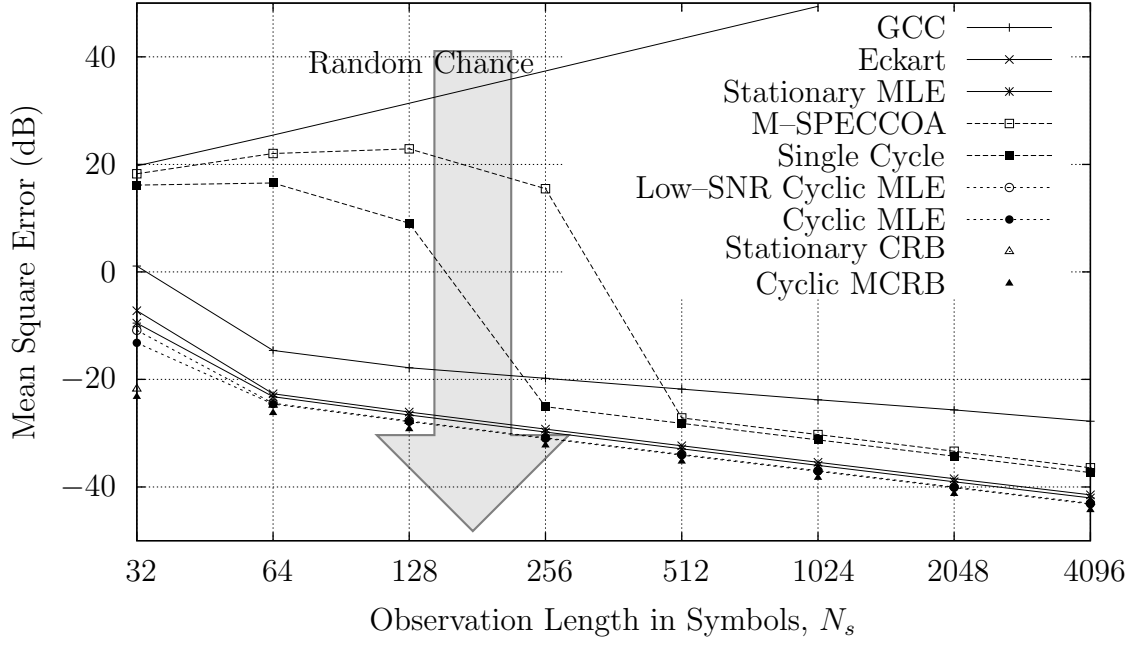
Figure 40.   TDOA Comparison with Quadratic Interpolation
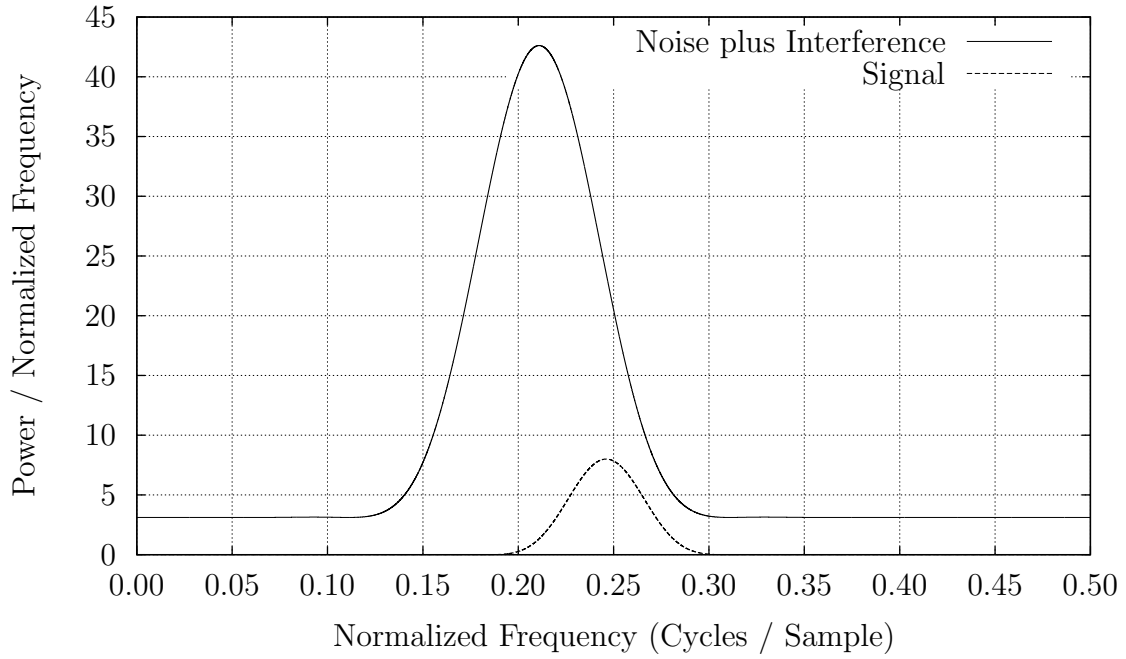
Figure 41.   Wide Band Interference Environment, showing the PSDs of both signal and noise plus interference
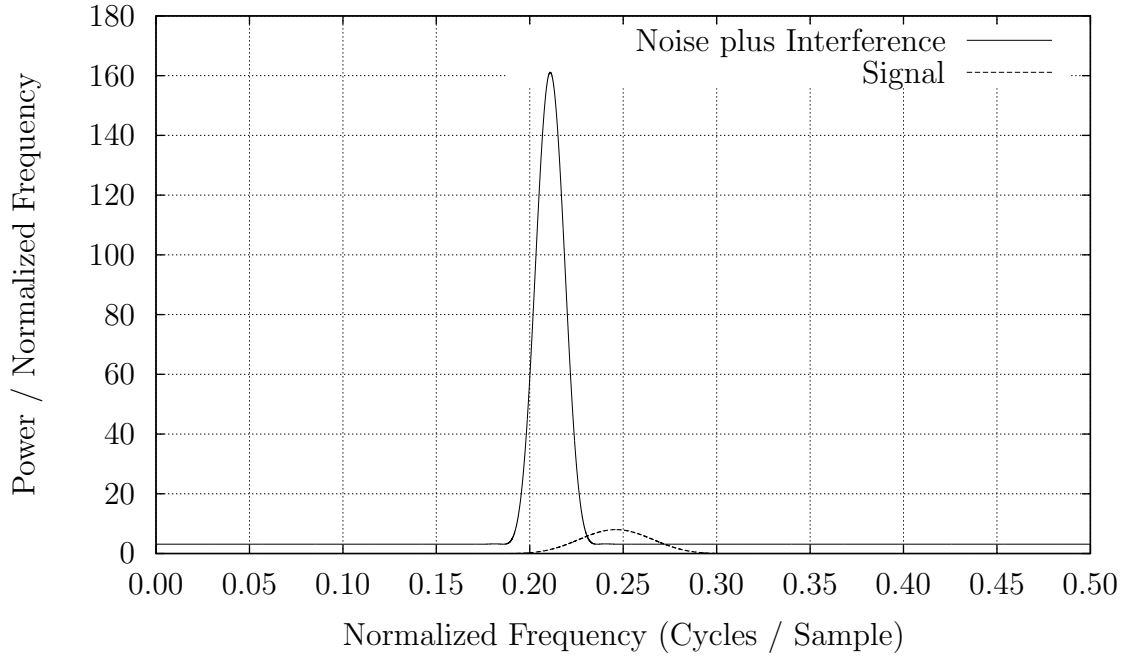
Figure 42.   Narrow Band Interference Environment, showing the PSDs of both signal and noise plus interference

by the noise plus interference. This is in general a bad thing, but here it is used to create *worst case* test environments.

Performance for the wideband environment is shown in Fig. 43 as a function of the observation length in symbols. As before, an MSE of 0 dB corresponds to a one symbol error in TDOA estimation, while an MSE of -24 dB corresponds to a one sample error input error. Further, since upsampling was used in this case, an MSE of -48 dB corresponds to a one sample error after upsampling. In general, from this plot, the ordering shown in Table 7 is apparent, suggesting some immediate conclusions.

The first conclusion is that the overwhitener is very important. Every algorithm having an overwhitener outperformed its counterpart without. This difference in performance can be as large as a factor of 40 dB, in the case of the cyclic MLE at an observation length of 2^9 symbols. Since applying an overwhitener is fairly cheap

153
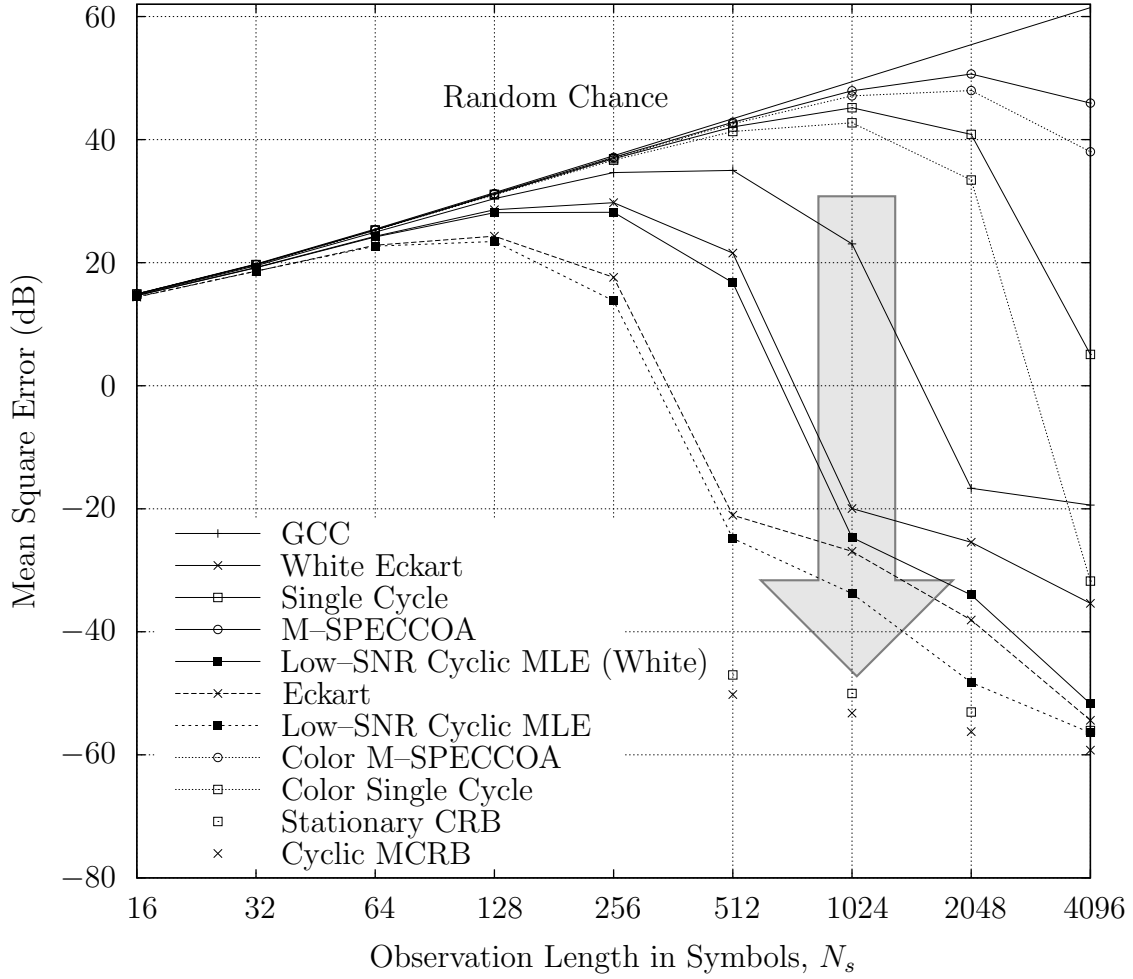
Figure 43.   Wide Band TDOA Performance

Table 7.   TDOA Estimator Ranking

Rank	Name	Overwhitener	Equation	Page
1.	Low–SNR Cyclic MLE	Yes	139	88
2.	Eckart	Yes	23	26
3.	Low–SNR Cyclic MLE (White)	No		
4.	White Eckart	No		
5.	GCC	No	21	25
6.	Color Single cycle	Yes	147	38
7.	Single cycle	No		
8.	Color M–SPECCOA	Yes	144	92
9.	M–SPECCOA	No		

154

computationally, this marks an easy improvement to any TDOA estimator working in a burst signal environment.

A second conclusion is that the Eckart filter outperforms every other TDOA estimation method shown here save the cyclic MLE. This suggests a couple of things. First, the single cycle TDOA estimation methods are no better than they were in the white noise case, even though they supposedly work the best in these strong interference environments. Second, it suggests that the portion of the MLE corresponding to the Eckart filter, $B_0[\tau_{d1}]$, is again the strongest component among all the other MLE components. Indeed, no cyclic algorithm performed better than the Eckart filter without using this component. That is to say, in order to do better than Eckart, his filter needs to be used as a starting point.

Third, the single cycle estimates, which are supposedly immune to interference, are hindered the most by that interference. This can be seen by the difference between M–SPECCOA and the color M–SPECCOA, and likewise between the single cycle and color single cycle estimator. Adding the overwhitener, as in the color versions of these algorithms, greatly improves their performance. This disproves the philosophy that single cycle TDOA estimators, for non–zero cycle frequencies, are capable of entirely removing the noise. Instead, mitigating the noise through filtering would appear to be a much better option.

A final observation can be made regarding the shape of the performance curves. This shape is distinctly different from the shape presented earlier on page 152 in Fig. 40. In particular, the estimation error starts out near the random chance line, curves down towards a bend, and then curves down towards a second bend. This is most obvious on the low–SNR cyclic MLE curve where it first curves towards 2^9 symbols, and then curves again towards 2^{11} symbols. This double set of curves can be directly attributed to the shape of the filtered cross correlation function between the two signals, as shown in Fig. 44. In particular, that filtered cross correlation has a peak at the true TDOA, $\tau_{d1} = 24$. When the signal is at bandpass frequencies,

155
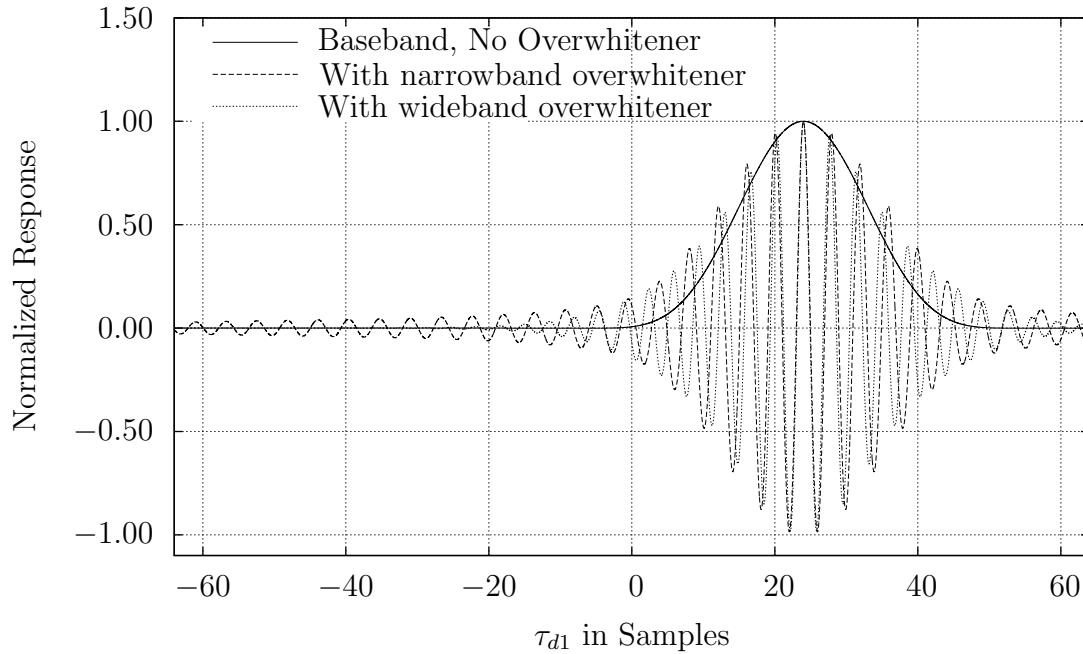
Figure 44.   Expected value at the output of the Eckart filter

such as were used in this test, this peak is surrounded by several local maxima near it. This explains the second curve and bend. The first curve is caused by spurious maxima being mistaken for the true peak. The second curve and bend is caused by local maxima near the true peak, that are mistaken for the true maximum. Once past this extra bend, the curves can be expected to flatten out much as before.

The next interference test estimates TDOA in the presence of a large narrow band interferer. The performance curves for this test are shown in Fig. 45. These performance curves require a change in the algorithm ranking noted earlier in Table 7. Two changes in particular are worth noting. The first is that the GCC algorithm performs much worse in this test case. Second, the color M–SPECCOA method now outperforms the single cycle estimator without the overwhitener. After noting these differences in ranking, some other differences are also immediately apparent.

Chief among these other differences is the fact that the cyclic MLE is no longer 15–20% better than the Eckart filter as it was in Fig. 37 in the last section.
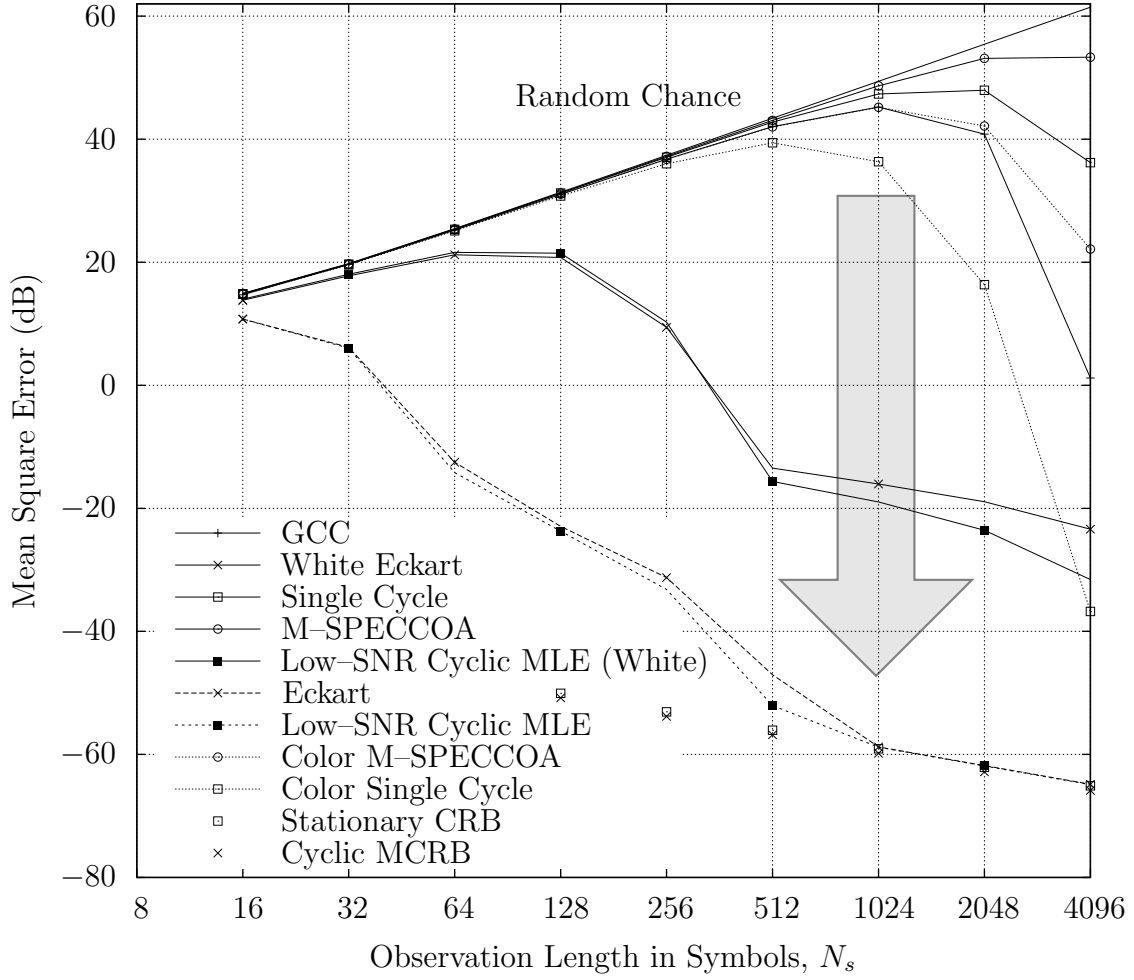
Figure 45. Narrow Band TDOA Performance

Similarly, the non–overwhitened versions of these two filters show less separation than before. This lack of separation can be traced to the fact that the interferer roughly covers the left half of the signal's spectra, from 0.184 CPS to 0.246 CPS (see Fig. 42). The problem is that extra components, $B_1, \ldots, B_4$ used in the cyclic MLE all compare the left half of the signals spectra, from 0.184 CPS to 0.246 CPS in this case, on one sensor with the right half, 0.246 CPS to 0.309 CPS, on the other. In this comparison, the narrowband interferer overwhelms the signal contained from 0.184 CPS to 0.246 CPS. While applying the overwhitener minimizes the impact of this interference by notching out the left half band, it also notches out the signal—effectively eliminating the contributions of $B_1, \ldots, B_4$.

Second, we see the same double bend as in the wideband case, only this time it is more pronounced. In both the cyclic MLE and the Eckart filter results, there are bends at 2^6 samples and 2^{10} samples. As before, this extra bend is caused by the fact that these signals are at bandpass frequencies.

Finally, the $B_0 [\tau_{d1}]$ component of the cyclic MLE still appears to contain the majority of the information available to a TDOA estimate. This supports the conclusion that while cyclic information can be used to improve TDOA estimation, it does not necessarily improve performance by orders of magnitude over optimal stationary methods—even in severe interference environments.

That returns the discussion back to how best to use cyclic spectral information. Such information is certainly signal specific, allowing an algorithm designer to create a signal specific estimator. Further, it is unique to the signal in that no other signal contains that information. However, it is not the strongest feature available to a TDOA estimation algorithm. The best information remains the stationary correlation when properly filtered, such as the Eckart filter. Added to this correlation, cyclic information may be used to provide a marginal improvement.

4.3.3 Unknown Scale TDOA Approximations. The final TDOA estimation test involves looking at the impact of the approximations used in the lopsided estimator of Eqn. (141) on page 89. As you may recall, these approximations were necessary to make the cyclic TDOA estimator practical in a real world environment without requiring the signal strength to be estimated. The lopsided estimator that resulted dropped one cyclic term and scaled the larger term in a non–optimal manner. This section will also present the impact of dropping this term from the MLE, as in Streight's version of the MLE.

To accomplish this test, we return to the baseband QPSK signal in white noise presented in Sec. 4.3.1, whose performance is shown in Fig. 37 on page 144. As with that test, only 128 symbols of signal will be collected. Further, all of the estimators that do not include the Eckart filter component, $B_0[\tau_{d1}]$, other than the GCC have been dropped from this test both for clarity, and because these other estimators performed so poorly in the last two sections.

That leaves four estimates under test. The first is the cyclic MLE, followed by the lopsided estimator, the Eckart filter, and then the GCC. Throughout the test, the SNR on sensor zero will be held constant at 4 dB, while the SNR on sensor one ranges from -14 to 4 dB. The results of this test are shown in Fig. 46. Fig. 47 also shows the same information, only it presents a zoomed–in view of the convergence region.

A couple of things are noteworthy from this chart. First, the difference between the lopsided estimator and the low–SNR cyclic MLE is the largest for the largest SNRs on sensor one. This is reasonable, since such SNRs violate the assumption underlying the lopsided estimator that the SNR on signal zero would be much larger than the SNR on signal one. Instead the SNR on sensor zero is nearly equivalent to the SNR on sensor one as it approaches 4 dB. A second noteworthy fact is that the lopsided estimator outperforms the Eckart filter everywhere making it a practical way of implementing a cyclic TDOA estimator. Third, the greater A_0 is over A_1, the

159
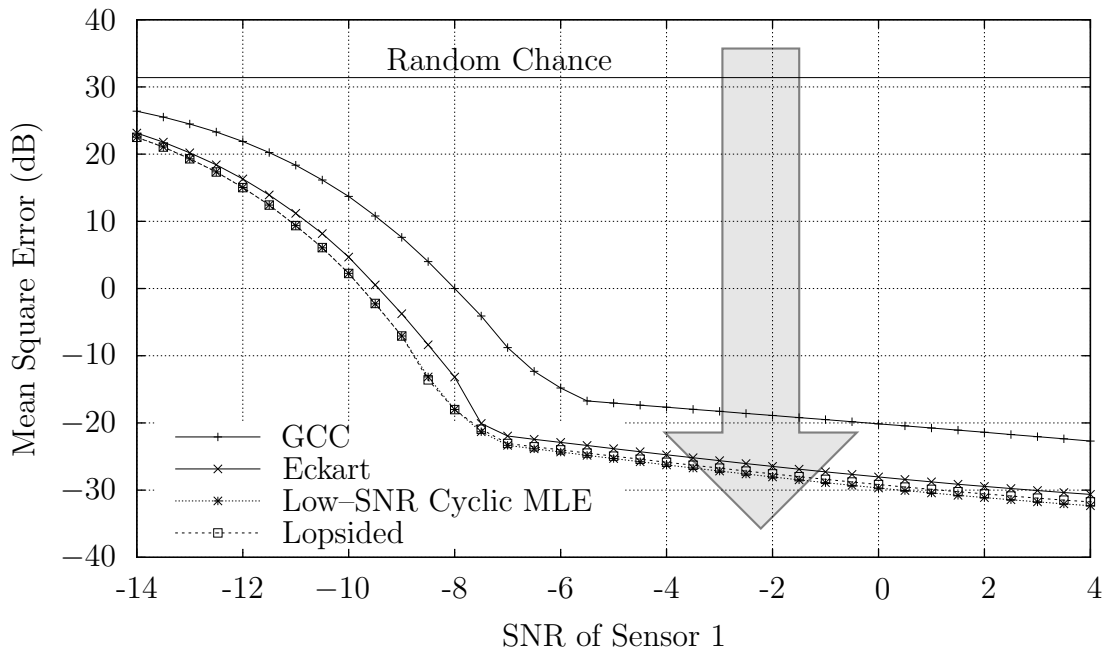
Figure 46.   Unknown Scale Approximation TDOA Estimation
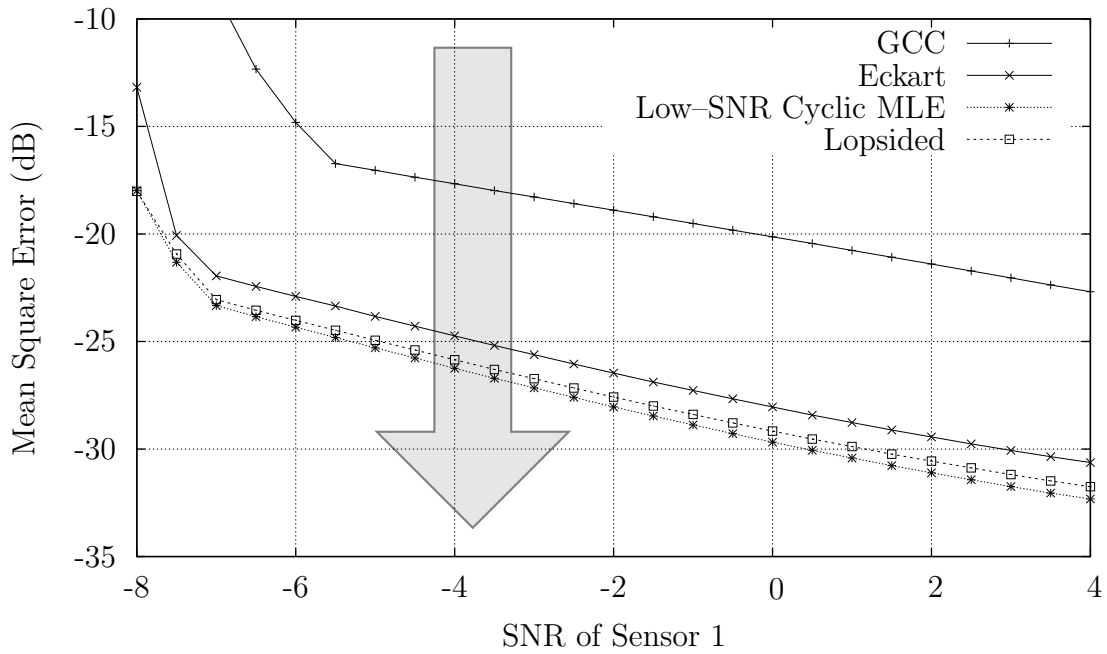
Figure 47.   Enlarged view of Fig. 46

160

closer this approximation comes to the cyclic MLE. All three of these observations suggest that this is a decent approximation to the cyclic MLE when the actual signal amplitudes are unknown.

Of all of the cyclic TDOA estimators, this estimator is certainly the most practical. It outperforms the Eckart filter, and it does not rely on extra knowledge such as the relative signal scales. However, the fact that this estimator does not outperform the Eckart filter by more than a small percentage calls into question whether or not this gain in TDOA estimation is worth the complexity of implementing it.

4.4   Presence Detection

This brings us to the last application area: signal presence detection. Having demonstrated the utility of this linear subspace approach to estimating parameters in burst communications systems, this section builds upon that foundation by testing several detectors built upon these similar estimates. In particular, three detectors were derived in Chapt. III. Two of those detectors, the multicycle ratio detector and the single cycle ratio detector, are new and novel. The other detector, the multicycle detector, is well known [17]. In this section, all three of these detectors will be tested via simulation. As these simulations will show, the new ratio detectors offer a degree of selectivity not found in any other burst signal detector.

To do this, four separate tests of these new cyclic ratio detectors are presented here. These simulations will first demonstrate their capability in general, and then ultimately their signal selectivity. The first test measures a detector's capability under known circumstances. The second test measures a detector's capability in a background of randomly changing Gaussian noise. Finally, the last two tests measure the discrimination capability of these detectors when broadband interference bursts are present, and second when similar interference bursts are present. The guiding question in each test is to ascertain whether or not the cyclic ratio detectors will achieve greater signal selectivity in a changing environment than any other detector.
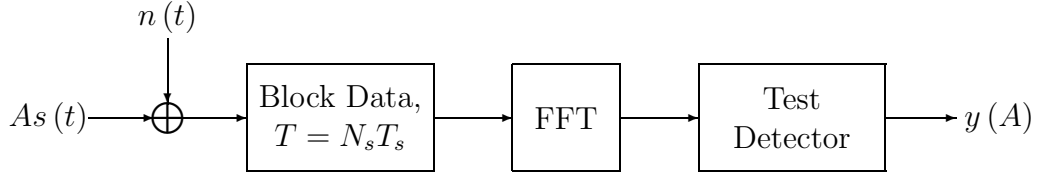
Figure 48.　Detection Test Setup

Each of these tests will need to measure what the probability is of properly detecting a signal, or accurately rejecting an interferer. This can be accomplished using the experimental setup shown in Fig. 48. In particular, the signal of interest, or in the last two tests an interference burst, together with noise will be grouped into blocks the size of the observation length. Following a Fourier transform, each of these blocks will be sent to the detector under test. In each case, the output of the detector is a function of the signal scale of the input signal or interference burst, A. From these outputs, the probability of correctly detecting the signal or incorrectly detecting on the interference burst can be measured by,

$$
P_{\mathrm{D}}(A) \approx \frac{1}{N} \sum_{n=0}^{N-1} \begin{cases} 1 & y_n(A) > \eta \\ 0 & y_n(A) \leq \eta, \end{cases} \tag{190}
$$

where $y_n(A)$ is the output of the detector for the nth test, and η is a threshold chosen to achieve some probability of false alarm,

$$
\eta = \left\{ \eta \middle| P_{\mathrm{D}}(0) = P_{\mathrm{FA}} \right\}. \tag{191}
$$

Using this approach, the first step to introducing these tests is still to introduce the detectors and then the signal under test. Nine detectors, listed below, will be tested and compared against each other. Of these nine, three of them will be single cycle detectors, while another three will be single cycle ratio detectors.

Radiometer A detector built off of the measurement of the energy in the observation alone, and given by Eqn. (33) on page 34. The bandwidth of this radiometer, W, has been set to $\frac{2}{T_s}$, the nominal null–to–null bandwidth of the signal.

Multicycle The optimal detector for detecting a weak cyclostationary signal in a Gaussian noise environment, given in Eqn. (168) on page 98 [17].

Zero–Cycle Also known as the Spectral Matching Detector [59], this detector is the optimal stationary detector [17]. It is created from the $\alpha = 0$ term of the multicycle detector.

Carrier Detector An approximation to the multicycle detector, using the feature that the signal is symmetric about the carrier frequency. (See Eqn. (37) on page 38.)

Symbol Rate Detector Similar to the carrier detector, except this detector uses only the redundancy found at intervals of a baud rate.

Multicycle Ratio The detector developed in Sec. 3.4.2 as the optimal detector for detecting weak digital communications signals in a Gaussian noise environment of unknown scale. This detector was presented in Eqn. (176) on page 101.

Zero–Cycle Ratio Similar to the multicycle ratio detector, except that only the $\alpha = 0$ terms are used.

Carrier Ratio Similar to the multicycle ratio, except only the carrier redundancy is used. (See Eqn. (178) on page 101.)

Symbol Rate Ratio A ratio detector created by dropping all but the symbol rate redundancy from the multicycle detector. (See Eqn. (177) on page 101.)

In order to apply all of these detectors, and in particular both the carrier and the symbol rate detectors and their ratio counterparts, the signal in question must be

Table 8.   Signal Parameters for the Detection Tests

Signal Parameter	Parameter Value
Modulation Type	BPSK
Symbol Rate, $\frac{1}{T_s}$	0.0625
Carrier Frequency, f_c	0.25
Number of Symbols, N_s	256
Pulse Function, $\psi(t)$	Rectangle

a BPSK signal.[9] The symbol length for this BPSK signal was chosen arbitrarily to be 16 samples, the carrier at 0.25 CPS, and then performance was compared for an observation lengths of 256 symbols. Further, the pulse shape in question was a bandlimited rectangle pulse, $\psi_R(t)$ (see App. A), chosen both for its simplicity and to match others who have tested similar signals previously [17]. As a final note, the first several tests were conducted in a white Gaussian noise environment such that $\mathbf{R}_n \propto \mathbf{I}$. These signal parameters are summarized in Table 8.

Under known conditions, the probability of correctly detecting the signal with only a 0.1% false alarm rate is shown in Fig. 49 as a function of the signal to noise ratio. The detector shown with the highest probability of detection at the lowest SNR is the best detector in this test. In particular, the best detector shown is the multicycle detector. As for the new ratio detectors, they each perform slightly worse (roughly 1 dB) than their non–ratio counterparts.

Each of these two observations can be understood by the conditions under which these detectors were created. The multicycle detector was designed to be the optimal detector in low SNR conditions. It should come as no surprise then that this is the optimal detector in this test case, since the test exactly mirrors the assumptions under which it was derived. Similarly two arguments can be made to explain why the ratio detectors fair worse. First, the difference between the ratio

---

[9]If the signal is a complex PAM signal, all but the Carrier detectors can be used.
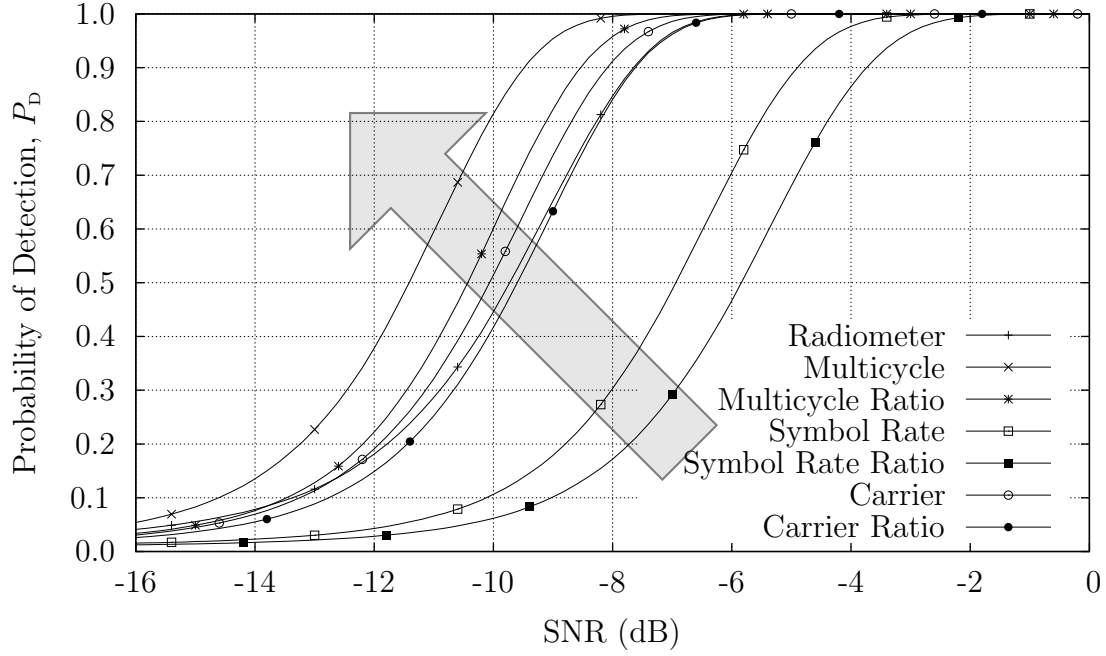
Figure 49.   Detection Capability when $P_{\mathrm{FA}} = 0.01$

detector and its non–ratio counterpart has an analogy in subspace detection. In subspace detection, an F test is a ratio test derived under similar hypotheses to those considered here, while a χ^2 test is its non–ratio counterpart. These two tests, the χ^2 test and the F test, are remarkably similar to the multicycle test and the multicycle ratio test respectively. They are so similar that comparisons are made between them in Appendix I will demonstrate little difference between their performance. Unlike the multicycle and multicycle ratio tests, however, comparisons between χ^2 tests and F tests are well known: the F test is known to perform worse than the χ^2 test [53]. Second, it only makes sense that to get the signal selectivity desired, some price must be paid for additional mistakes. Such a mistake might be determining that a signal is present, but that it is not the signal of interest. Standard cyclic detectors are not designed to avoid this type of mistake, hence they have no burst differentiation capability either.

Probability of Detection, P_D

SNR (dB)

Radiometer
Multicycle
Multicycle Ratio
Zero Cycle
Zero Cycle Ratio
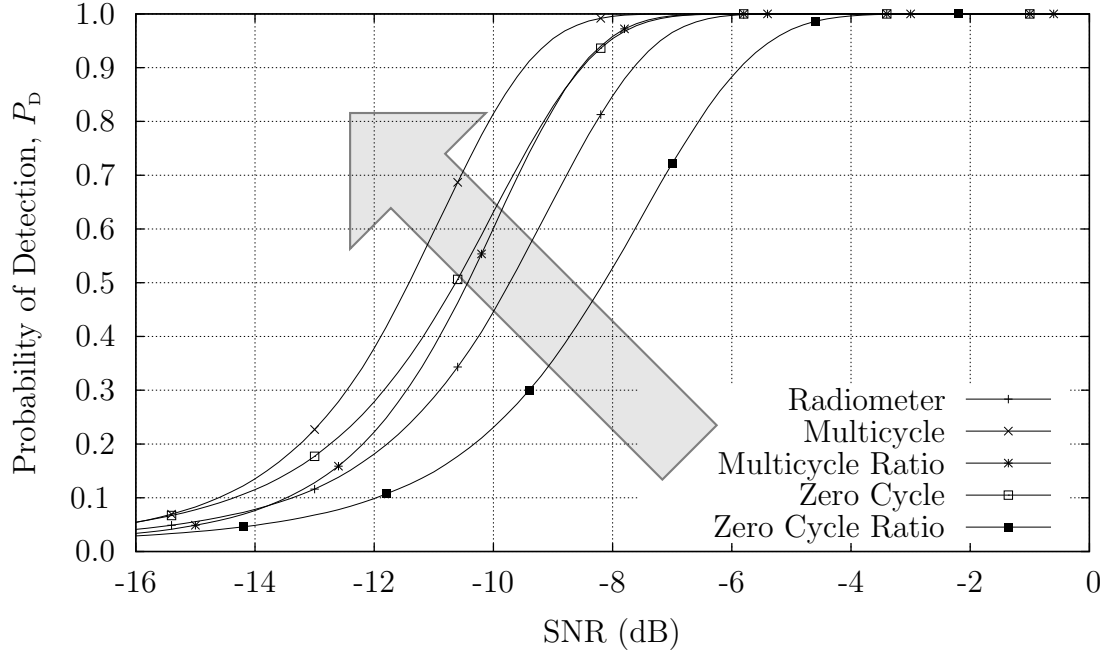
Figure 50.   Stationary vs. Cyclostationary Detector Capability, $P_\mathrm{FA} = 0.01$

An interesting comparison to make here is to look at the difference between the optimal stationary signal detector, known as the zero–cycle detector, and the optimal cyclostationary signal detector or multicycle detector. Given that these detectors are the best stationary and cyclostationary detectors, this comparison should help to illustrate the performance benefits between these approaches to describing digital communications signals. This comparison is shown in Fig. 50. Among the non–ratio detectors, the multicycle detector shows only about a 1 dB improvement in detection capability over the zero–cycle detector. Among ratio detectors, the improvement increases to a 2 dB gain achieved by using cyclic methods over stationary ones.

That brings us to the second test case, where the performance of each of these detectors is measured in a white noise background environment with a randomly changing scale. Such an environment might be caused by either a very wideband interferer or equivalently an automatic gain control system attempting to compensate for energy outside of the band of interest. The new ratio detectors should perform

quite well under these conditions, since they were created under the assumption of an unknown or changing background scale.

The particular test presented here was introduced by Gardner as a reason for using cyclic detectors over standard energy detectors [17]. In his paper, Gardner demonstrates that a single cycle detector outperforms a radiometer in a changing noise environment. The changing noise environment is created from a white noise sequence created such that $\mathbf{R}_n = 2\sigma^2\mathbf{I}$. The variance parameter, σ^2, is itself generated once per test by,

$$\sigma = \sqrt{\frac{|z|}{2}}, \tag{192}$$

where z is selected randomly from a Gaussian probability distribution such that the mean of z produces an SNR of -10 dB, and the variance of z is one tenth its mean.[10] The only difference from this test and Gardner's test before it is the setting of the mean of z. Gardner's test was accomplished for a mean that produced zero SNR conditions, while the mean of z has been adjusted here in order to make the results stand out more.

Figure 51 shows the probability of detection for each of these tests, compared to its probability of false alarm. If you compare the results for the non–ratio tests with the results in [17], you will find them to be essentially identical—confirming that this test validly represents both the detectors and the conditions presented there. Fig. 51, however, adds the performance of the cyclic ratio detectors to Gardner's original results. According to this new figure, the multicycle ratio detector outperforms every test on the chart—having the highest probability of detection for any given false alarm rate. This is consistent with the conditions under which it was derived: it is the optimal detector for detecting a signal in an unknown or changing background

---

[10]While it might make more sense to select z from a probability distribution defined for positive numbers only, this choice has been made to mimic the performance test in [17].
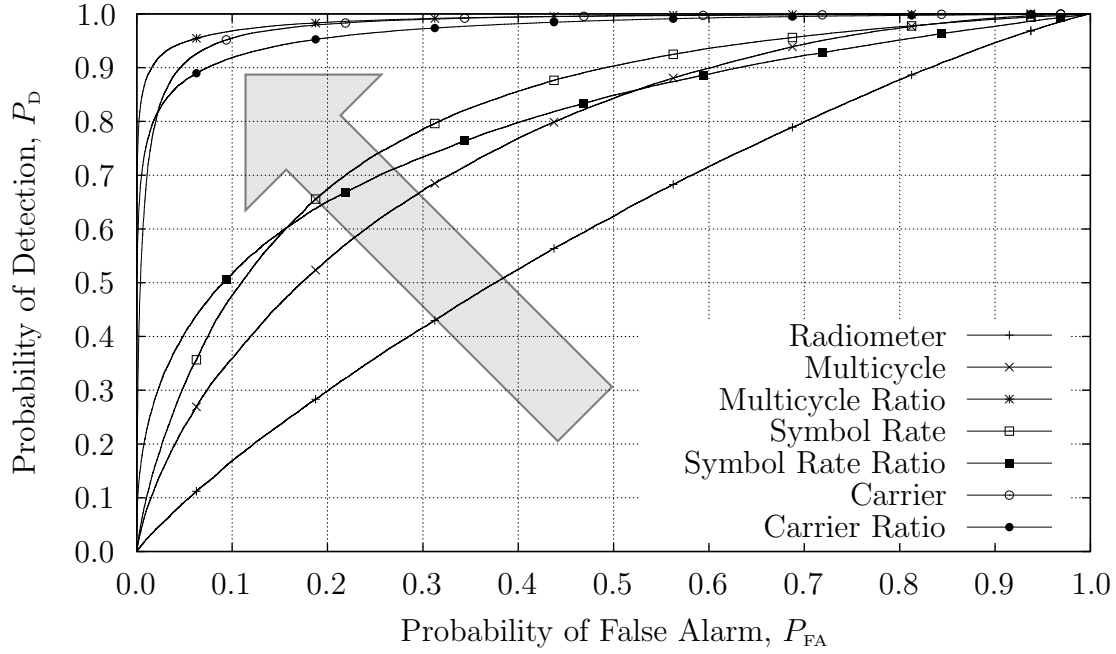
Figure 51.   A Repeat of Gardner's Test

environment. A second feature worth noting is that each of the single cycle ratio detectors outperforms its non–ratio counterpart when the probability of false alarm is low. When the false alarm rate is much higher, the non–ratio tests perform better. This performance difference can be directly traced to the randomly changing noise level. Once the standard single cycle detectors suffer false alarms from all of the tests where the random noise level was relatively high, they then outperform their ratio counterparts.

The zero cycle component in this test, shown in Fig. 52, is particularly revealing. First, the optimal stationary ratio detector, the zero–cycle ratio, performs fairly well in this test. Indeed, it is better than all but the multicycle ratio and carrier based tests. Had the comparison been made between cyclostationary detectors and the radiometer alone, leaving out the optimal stationary detectors, such results would have presented a misleading picture of the capability of stationary detectors. The reason for this is the same as the reason why Eckart's filter did so well at TDOA

168
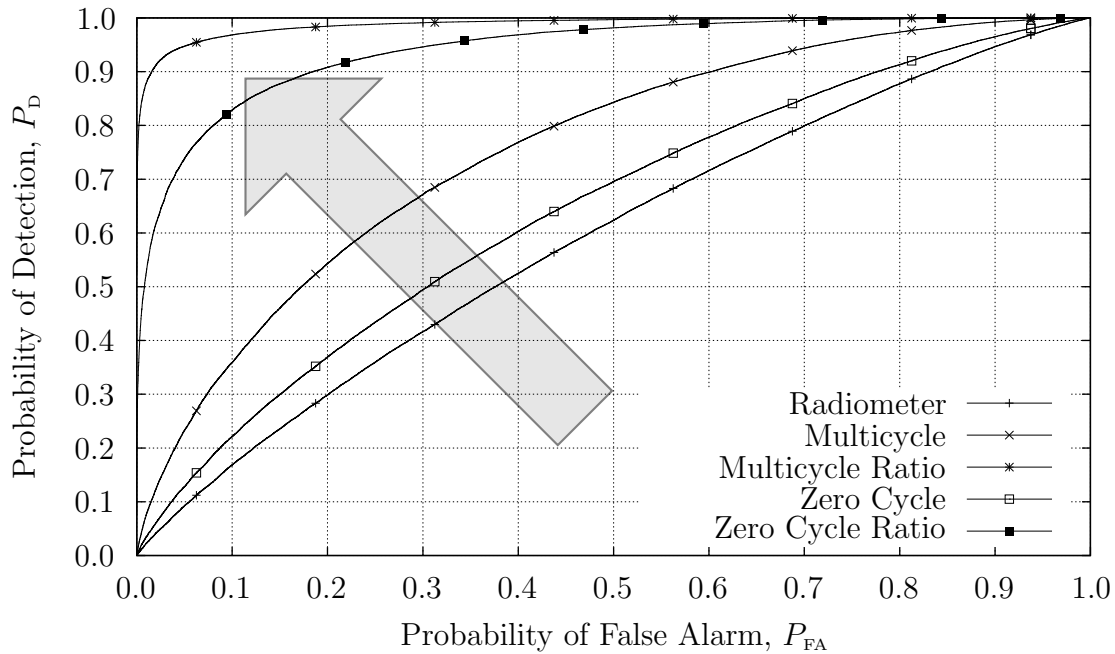
Figure 52.   Optimal Detector Performance in Gardner's Test

estimation—stationary or zero–cycle statistics are much stronger than the baud rate and symbol rate statistics used by these other detectors.

That brings us to the third test, the test of changing background levels. In this test, the signal is not present at all. Instead, a white interference burst of varying strength will be present in each detection interval. Under these conditions all detections are false alarms, since nothing but interference is present. Fig. 53 shows the results of this test. As one might expect, the ratio tests are completely resistant to white interference bursts. All of the non–ratio detectors, however, perform poorly under these conditions. That is because these tests are quadratic tests of the input. If the input gets a sudden boost in noise, the detection statistic responds with the square of the boost. This renders the threshold, which was set in order to create a 0.1% false alarm rate on the training data, inappropriate whenever such burst interference is present. The ratio detectors, because they take a ratio of one quadratic test to another, are immune to this change. Thus the scale of the noise can change to
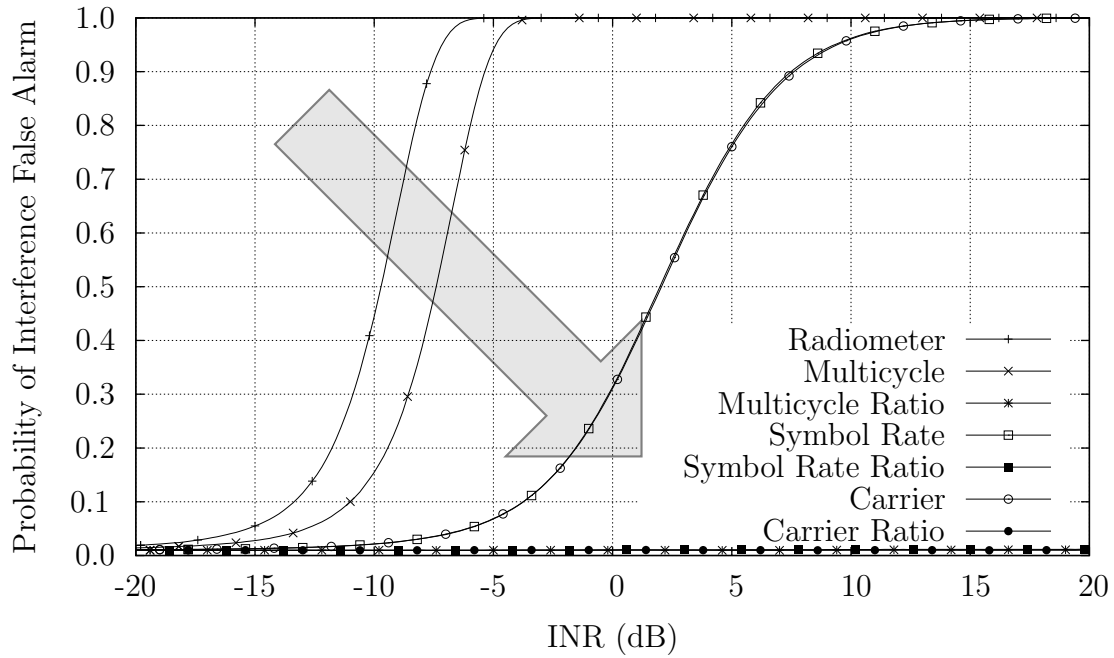
Figure 53.   Wideband Interference Resistance in Cyclic Detectors

anything, and the false alarm rates of these detectors will remain the same. This is another way of saying that these detectors have a constant false alarm rate (CFAR) in a changing noise environment—exactly what Theorem 2 proved.

Fig. 54 compares the interference resistance found in the optimal stationary detectors with the same resistance found in the optimal cyclostationary detectors. From this figure, as with Fig. 53 previously, the strength of the ratio detectors is obvious. Unlike the radiometer or the non–ratio detectors, both of the ratio detectors in this figure are completely resistant to changes in the background noise level.

These two tests confirm that the total energy measurement, found in the denominator of the ratio detectors,[11] does indeed track the noise level as it fluctuates, as predicted by the discussion in Sec. 3.4.2. By entirely removing this noise level uncertainty, Figs. 53 and 54 show that these new detectors no longer suffer from the limitations of non–ratio detectors.

---

[11]See, for example, Eqns. (176), (177), and (178).
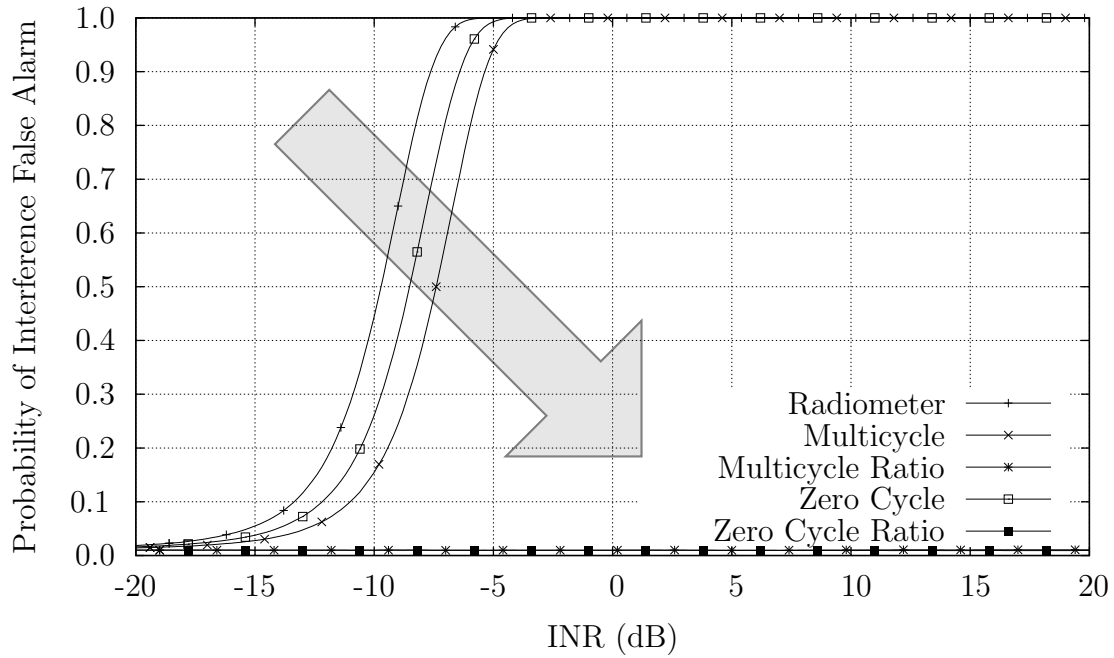
Figure 54.   Wideband Interference Resistance in Optimal Detectors

Table 9.   Similar Interference Parameters

	Carrier Rate, f_c	Symbol Rate, $\frac{1}{T_s}$
Signal of Interest	0.25000 CPS	0.06250 CPS
Similar Interferer	0.25024 CPS	0.06274 CPS

A more interesting test would show how well this detector discriminates against interference bursts that are similar in structure to the signal burst itself. This is the last of the four tests and practically it is the most important. Thus, the last test measures how a given detector responds to a burst having a slightly different carrier, and a slightly different baud rate. In particular, the interferer is given a carrier of $\frac{1025}{4096} \approx 0.25024$ CPS and a baud rate of $\frac{257}{4096} \approx 0.06274$ CPS. See Table 9 to compare these parameters with the signal of interest. These values were chosen since they are one FFT bin $\left(\frac{1}{4096} \text{ CPS} \right)$ off from the true values (0.25 and 0.0625 CPS). Performance under these conditions is shown in Fig. 55. As with the last test, no signal is present in this test. Therefore any response from a detector in this environment represents
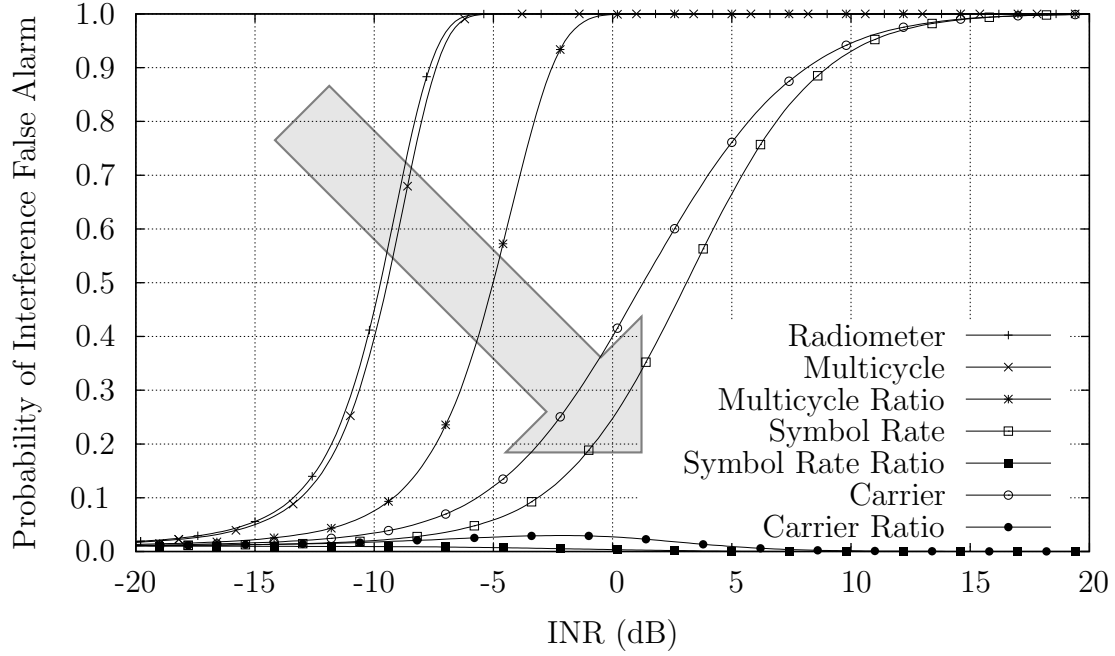
Figure 55. Resistance to Near–Identical Burst Interference, $P_{\mathrm{FA}} = 0.01$

an increase in the false alarm rate of the detector, or equivalently a burst interferer that was incorrectly identified as the signal of interest. The results shown in Fig. 55 are similar to those of the wideband band burst interference test case. The obvious conclusion is that standard cyclostationary detectors offer only minimal capability of rejecting burst interference.

Unlike the wideband interference test, however, the multicycle ratio detector does not discriminate against these similar interferers. The reason for this can be traced to the zero cycle component of the multicycle detector, which exploits the PSD of the signal to detect it. However, in this case, the PSD of the interferer was nearly identical to the PSD of the signal (see Fig. 56). The zero cycle component of the multicycle detector attempts to use this PSD to determine whether or not a signal was present. Since the interfering signal has nearly the same PSD, any examination of the PSD strongly suggests that the signal of interest is present.
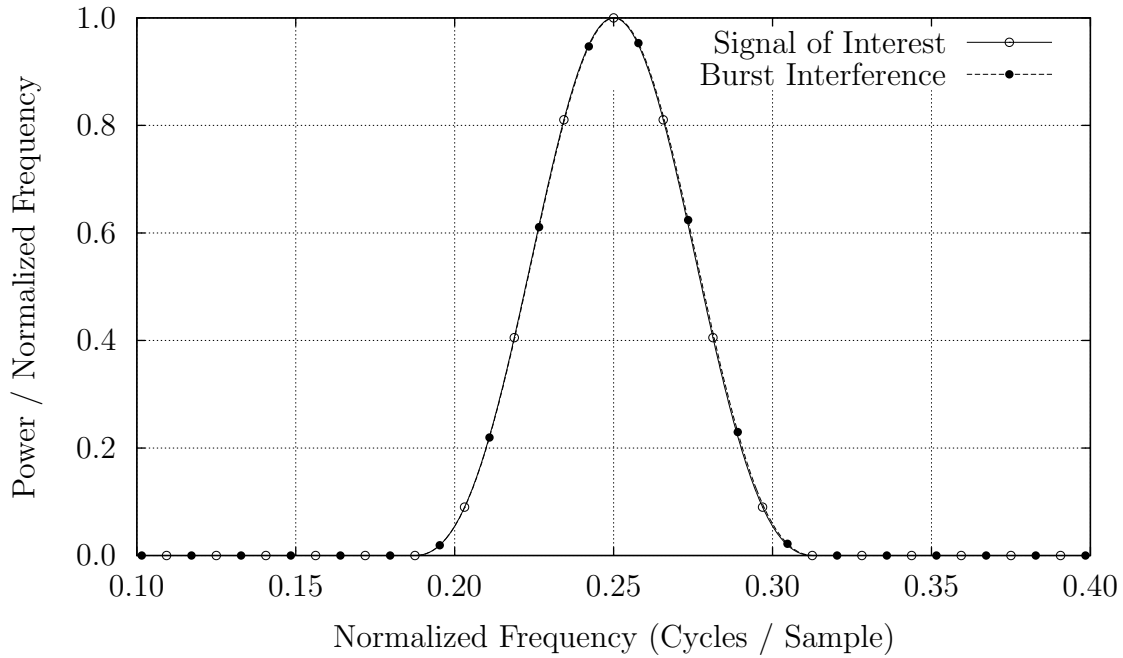
Figure 56.   PSD of Signal, and Similar Interferer

This makes one test case where the stationary component of the signal, containing the strongest piece of information, is deceptive, since every detector that uses it here performs poorly. Perhaps this would be clearer on a chart showing only those detectors using the zero cycle term, such as Fig. 57, where the zero–cycle component of the multicycle detector is separated from the other components for comparison. By using the PSD of the signal, this zero–cycle component has the highest probability of alarming on the similar interference burst. Likewise, it has the worst resistance among all of the ratio detectors. Although adding the other cyclic components of the multicycle detector to the zero–cycle detector improves its resistance to burst interference, it does not improve it much.

What the multicycle detector lacks in this similar burst interference environment, the single cycle ratio detectors pick up. The two single cycle detectors, the carrier ratio and symbol rate ratio detectors, are the clear winners in this test. While neither of these two detectors are the "best" under known conditions, such as Fig. 49,

Probability of Interference False Alarm

INR (dB)

Radiometer
Multicycle
Multicycle Ratio
Zero Cycle
Zero Cycle Ratio
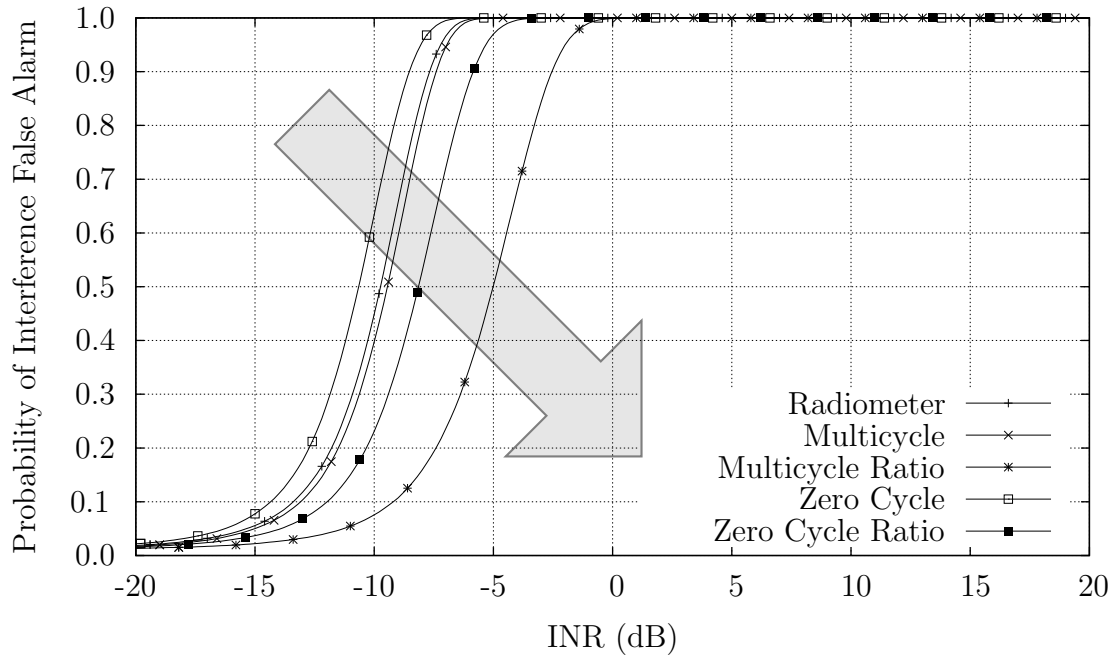
Figure 57.   Zero–Cycle Resistance to Near–Identical Burst Interference, $P_{\mathrm{FA}} = 0.01$

none of the other detectors match the signal selectivity found in these two detectors
otherwise.

All told, these tests contrast the strengths of the ratio detectors against the
limitations of previous cyclostationary detectors. While it is true, mathematically,
that estimating a spectral correlation over a long period of time completely removes
the noise, that does not mean that such detectors are immune to noise after only
finite periods of time. A strong burst of interference will always trip a non–ratio
cyclostationary detector—regardless of whether or not that burst of interference has
the feature of interest or not. While mathematically such detectors may be immune
to all interfering signals as the observation length becomes infinite, practically none
of the standard cyclic detectors are immune to interference bursts at all. Instead,
only the cyclic ratio detectors show true immunity to burst interference.

V. Conclusions

Given that the initial goal of this research was to build a better burst signal detector for high interference environments, this research has accomplished that and more. By returning to first principles, not only were interference resistant burst detectors demonstrated, but also Minimum Mean Square Error (MMSE) single and dual sensor filters, demodulator performance prediction algorithms, and maximum likelihood interference–resistant TDOA estimators were demonstrated as well. In order to summarize these results, this chapter is split into two sections. The first section will present a quick summary of the capability of the algorithms developed using the linear subspace approach. Then, motivated by the success of the algorithms developed here, the second section will discuss some additional application areas and recommend topics for future research.

5.1 Summary of Findings

Tracing these new algorithms from their development through simulation requires looking at not only the new linear subspace approach, but at each of the three application areas central to this research. This summary will follow that pattern by first outlining the linear subspace approach, and then highlighting the consequences this approach had on each of the application areas studied in this research.

The linear subspace approach followed from some simple assumptions concerning the signal of interest and led to the application of classical statistical principles to fundamental communications signal processing problems. The three fundamental assumptions were that the signal had a Pulse Amplitude Modulation (PAM) structure, was finite in time, and that the pulse function underlying that PAM structure was known. Then, since the signal was finite in length, it became possible to allow the power spectral density of the noise and interference to be known, and not necessarily white, since it could be estimated. Further, since the signal had a PAM

structure, the signal was shown to lie within a subspace of all possible received waveforms. This subspace was traced through a Fourier transform, demonstrating that the signal's spectral content could be constructed by linear operations upon the underlying z–transform of the signal. Then a probability distribution, that the underlying z–transform had a Gaussian distribution, was proposed for the signal. This distribution was then validated in Sec. 4.1—allieviating a problem which has plagued cyclostationary signal processing development for years [23]. As a result, classical statistical techniques such as maximum likelihood estimators, likelihood ratio detectors, and Cramér–Rao bounds, became a simple matter of calculation. This simple beginning then led to improved signal processing algorithms, which became the topic of the rest of this research.

Since this linear subspace approach preserved the first and second moments of the signal of interest and the interference, properties central to previous cyclostationary developments, tracing the algorithms resulting from this approach will reveal that several methods did not change, as outlined in Table 10. Applying the linear subspace approach under these previously studied assumptions, such as stationarity or white noise, led to standard algorithms. Yet the methods that have changed did so as a result of a more appropriate description of both the signal and the interference. Further, the new methods that resulted are much more resistant to interference than their predecessors. The rest of this summary section, therefore, follows the outline given in Table 10 by highlighting first the algorithms that have been preserved under this new model, followed by discussing each of the algorithms that have changed.

Starting from the beginning, Chapt. III derived several new algorithms from the simple assumptions that the signal of interest was a finite, Gaussian, Pulse Amplitude Modulated (PAM) waveform corrupted by additive Gaussian interference plus noise. Although this method did not assume that the transmitted signal was infinite, many of the resulting algorithms are similar or even identical to known cyclostationary algorithms derived under infinite signal conditions. Among the al-

Table 10.   Algorithms Resulting from the Linear Subspace Approach

Application	Assumptions	Algorithm	
Demodulation	Weak signal		Matched filter
	Complex signal		Berger & Tufts filter
	Real signal	*New*	BPSK MMSE filter
	Arbitrary filter	*New*	MSE Performance Prediction
	Many sensors	*New*	Multisensor BPSK MMSE filter
TDOA Estimation	Stationary signal, colored noise		Eckart's filter
	Cyclostationary signal, white noise	*New*	Implementable Cyclic MLE
	Cyclostationary signal, colored noise	*New*	Overwhitened Cyclic MLE
		New	Cramér–Rao bound
Detection	White noise		Gardner's Multicycle detector
	Colored noise		Rostaing's modifications
	Burst interference	*New*	Cyclic Ratio Detector

gorithms that have not changed are the matched filter for colored noise, Berger and Tufts' MMSE filter for baseband signals [2], and Gardner's optimal detector for weak cyclostationary signals in known noise environments [17]. Streight's ideal maximum likelihood TDOA estimator for white noise environments [61], on the other hand, was modified to include a missing term.[1]

The fact that these algorithms have not changed is not necessarily a failure of the linear subspace approach, rather this fact commends the validity of this approach. Given that some of these solutions remain unchanged after 35 years, they have certainly stood a test of time and are not likely to be improved upon. Therefore, the fact that the linear subspace approach leads to these already known and optimal methods suggests that, when applied under different assumptions, it will also lead to optimal methods.

Further, unlike previous attempts at deriving these cyclostationary algorithms, the linear subspace approach was based upon classical statistical principles. This was made possible by the introduction of an approximate probability density function to describe the signal in frequency. From this probability density function, cyclostationary properties in the signal became *consequences* of the underlying probability distribution, instead of an incomplete *definition* of the same underlying probability distribution. Further, even though this probability density is only approximate, tests presented in Sec. 4.1 demonstrated its validity for bursts as short as $N_s = 8$ symbols. Everything else in this dissertation then followed from this probability density function via classical statistical principles such as maximum likelihood estimators, Cramér–Rao bounds, and likelihood ratio tests.

Given a probability distribution function, the first task was to estimate the unknown message data. The reason for this is quite simple: an estimate of the message is required in order to create TDOA estimation and signal detection algorithms. Therefore, this linear subspace method was applied to demodulation in order to de-

---

[1]This is the B_3 shown in Eqn. (134) on page 87.

rive an optimal estimate of the underlying bits. Two filters were developed as a result of this process. The first, equivalent to Berger and Tufts' filter, was already known to be optimal for signals with a complex baseband. The second filter, a MMSE filter for BPSK signals, was a new result.

This BPSK MMSE filter was then tested in a laboratory simulation environment, in Sec. 4.2.1, and shown to outperform all other linear filters. This was true regardless of the interference environment or the multipath distortion suffered by the signal (Secs. 4.2.1 and 4.2.2) [26]. Further, the equalizer predicted by this method was shown to be identical to a similar equalizer achieved under adaptive conditions (Fig. 21). This latter fact, combined with Ericson's proof that all optimal demodulator filters can be factored into a matched filter and equalizer [12], together with the proof that adaptive equalizers truly achieve this optimal performance [30], provides additional validation that this filter is truly the minimum mean square error linear filter for digital communication signals.

Yet the development and analysis of optimal data estimators did not stop with the optimal single sensor BPSK linear filter. Three other consequences were shown from it as well. First, the same development led to performance prediction methods appropriate for arbitrary channel conditions and arbitrary BPSK demodulators (Sec. 3.2.2) [27]. Sec. 4.2.2 then validated these prediction methods in both colored noise and multipath interference environments. Second, Modified Cramér–Rao bounds were presented that were applicable to BPSK demodulator performance. These bounds, which were difficult to calculate under cyclostationary developments, were simple consequences of the linear subspace framework. Finally, the development presented the optimal form of a multi–sensor BPSK linear filter (Sec. 3.2.4). During performance simulations, this filter achieved 3.4 dB improvement (Fig. 35) over the single sensor filter in a simulated high interference environment.

Estimation, however, is only the first task. Judging from the performance just mentioned, linear subspace theory is certainly appropriate for generating estimates of

the transmitted data sequence. These estimates of the data sequence are then central to TDOA estimation and signal detection, as the algorithms for these two application areas depend upon these basic estimates. Therefore, we turn our attention to these next two application areas.

The next application area following BPSK demodulation, TDOA estimation, has historically been difficult for cyclostationary signal processing. Without a valid probability density function, it has been impossible to use classical statistical techniques to develop optimal estimators [23]. Indeed, the first maximum likelihood estimator, presented seven years after the initial cyclic TDOA developments by Streight in [61], was limited to low–SNR signals in benign interference environments. Using the linear subspace approach, however, things changed.

The first change to TDOA estimation was a new maximum likelihood TDOA estimator appropriate for all SNRs and all interference environments. Unlike Streight's ideal cyclic MLE [61], this estimator can be implemented. It was then tested in Secs. 4.3.1 and 4.3.2, marking the first open comparison of a cyclostationary TDOA estimator against either the Eckart filter or the Hannan–Thompson TDOA estimator. This testing demonstrated that cyclostationary TDOA estimation is indeed more powerful than these optimal stationary TDOA estimators, although the improvement is limited to about 15–20%. More importantly, however, this testing demonstrated the importance of using an overwhitener to limit the impact of the interference prior to TDOA estimation. This overwhitening technique has not been applied to cyclostationary TDOA estimation before, primarily because of the difficulty in deriving cyclostationary algorithms. Yet it is a natural consequence of the linear subspace approach.

The second change to TDOA estimation involved the re–examination of single cycle TDOA estimation methods. These methods were proposed by Gardner and Chen in [23] for improved TDOA estimation results in high interference environments. According to Gardner and Chen, these methods should be completely

resistant to interference after sufficient integration times. However, as Sec. 4.3.2 demonstrated, this fundamental assumption of interference removal is flawed. Stationary estimators, containing precisely the corrupted statistics that Gardner and Chen wished to avoid, outperformed every single cycle TDOA estimator. Thus, although these single cycle methods are valid, they are not very efficient.

The last change in TDOA estimation theory revolved around deriving Cramér–Rao bounds for estimating the TDOA of a digital communication system received by two separate sensors. Previous attempts at creating Cramér–Rao bounds appropriate for digital communication signals exhibiting cyclostationary properties had failed to account for the non–uniform power spectral densities common in high interference environments and the cost of estimating the necessary nuisance parameters. These new bounds solve both of these problems, making them valid and appropriate for estimating TDOA performance in arbitrary interference environments.

Comparing the performance of all of these estimators in benign and in high interference environments revealed several features not yet documented for cyclostationary TDOA estimators. In particular, the importance of subsample interpolation, a topic previously ignored in cyclostationary TDOA estimation, was demonstrated in TDOA error measurement. For example, Fig. 39 on page 148, which was generated without subsample interpolation, might falsely lead one to believe that the performance of the TDOA estimators in it were orders of magnitude better than their true performance, shown in Fig. 40 on page 152 which uses interpolation. In the past, this failure to do subsample interpolation has resulted in misleading results regarding cyclostationary estimator performance [24, 61]. Second, comparing the performance of these estimators in colored noise conditions demonstrated that the previous Cramér–Rao bound for stationary signals remains a valid estimate of the performance a cyclostationary algorithm might achieve. This contradicts performance shown in [24] and [61], where the subsample interpolation problem, together

181

with an overly simplified Cramér–Rao bound, combined to generate results that appeared orders of magnitude better than the Cramér–Rao bound on estimation error.

Following TDOA estimation, we finally returned to the problem that initially motivated this research: signal presence detection. Prior to this research, two classes of detection algorithms existed for burst signals. The first class detected signals based upon their energy. Detectors of this variety are well known for being susceptible to noise fluctuations [60], or equivalently to burst interference. The second class of detector, cyclostationary feature detectors, is well known for its signal selectivity, but its performance in a burst interference environment suffers in a similar manner to the energy detection methods (see Fig. 53). To overcome this problem of selectivity in the presence of burst signal interference, a new class of signal selective detector, the *cyclic ratio detector*, was developed using the linear subspace approach (Sec. 3.4.2). This detector is similar to previous cyclostationary detectors, save that it is normalized by the total energy within the band. This extra normalization, similar to the normalization found in an F–test [53], is the key to the interference resistance that this detector achieves (see the discussion concluding Sec. 3.4.2).

Then, in a simulation environment, this detector was compared against both energy detectors and cyclostationary detectors. When the interference environment was constant, these ratio detectors performed about 1 dB worse than the standard cyclostationary detectors (Fig. 49).

When these same detectors were tested in a burst interference environment, the new *cyclic ratio detectors* outperformed every other burst signal detector (Figs. 53 and 55). Each of the other detectors performed as expected in this environment: the energy detector tripped on the energy of the interference, and the cyclostationary detectors, after offering more resistance to the interference, eventually tripped on the strong interference bursts as well. Only the *cyclic ratio detectors* seemed immune to this interference. Indeed, as the interference increased, the baud–rate ratio and the

182

carrier rate ratio detectors saw a *decrease* in the number of false alarms, rather than the increase seen by every other detector.

Achieving this interference resistance was the ultimate goal and purpose of this work. The fact that it was achievable using a linear subspace approach certainly commends the utility of this new approach. Looking back over the simulation results, the performance of the methods derived using this approach was always as good as or better than the performance of previous methods. There was never a time when an optimal method derived using this approach performed worse.

5.2   Recommendations for Future Study

This brings us to the question of future research, given that this linear subspace method is now fairly well validated for digital communication signals. All that remains is to derive and demonstrate further algorithms using this model. While there is room for additional development within each of the three application areas, as discussed below, new algorithms need not be limited to these three applications.

In terms of demodulation, much more work is left to be done. First, simple modifications to the work presented here could be made to apply this work to other signaling types such as offset QPSK (OQPSK),[2] minimum shift keying (MSK), or binary coherent phase modulation (BCPM).[3] While more complicated, this work could also be extended to Frequency Shift Keying (FSK), and Orthogonal Frequency Division Multiplexing (OFDM). Second, this model can be used to develop optimal carrier and symbol synchronization algorithms. Yet such improvements do nothing to improve upon the biggest problems associated with this filtering technique: the unknowns. Thus, further filtering research should look into the problems of estimating the unknown pulse function, $\psi(t)$, the signal strength A, and the noise

---

[2]See App. C.

[3]See App. D.

PSD. Without solving these problems, the filters presented here cannot be applied to non–burst communications signals.

Further solutions can also be implemented in the area of TDOA estimation. First, Streight's analysis and development derived not only the maximum likelihood TDOA estimator, but the maximum likelihood joint TDOA and Frequency Difference of Arrival (FDOA) estimator [61]. By leveraging off of his example, extending this work to FDOA estimation in high interference environments should be fairly straightforward. Indeed, deriving a maximum likelihood TDOA/FDOA estimator shouldn't be much more difficult than the TDOA estimator that was presented here. What remains is to demonstrate this estimator and the gains associated with it. In particular, if Streight's tests were to be repeated, how much performance can be gained by knowing the signal pulse function and noise PSD? Third, these techniques make it possible to derive the optimal estimators in correlated, not just colored, noise environments—such as when multiple sensors see the same interfering signal. The particular assumption, that the interference is uncorrelated between sensors, is perhaps one of the least realistic assumptions made. Yet the difficulty of TDOA estimation in correlated noise environments has led others to make this same assumption [37]. Resolving this difficulty should follow from the linear subspace theory as well.

Finally, in the area of detection, one critical and fundamental problem remains: what is the probability of detecting a signal with any of the detectors discussed here? While the probability of detecting a signal with an energy detector is well known, that same probability is not as well understood for cyclostationary detectors—much less cyclic ratio detectors. A similar and related problem is the question of what threshold should be chosen to achieve a particular probability of a false alarm? Without knowing these probabilities, designing a communications system to avoid cyclostationary detection is an ill–defined problem. Likewise, designing a detection

184

system to operate in a high interference environment is a difficult problem without knowing the proper thresholds.

While this dissertation has presented three application areas, others could easily be mentioned here as well. In particular, Angle of Arrival (AoA) or Direction of Arrival (DoA) estimation algorithms could be improved by these techniques. While cyclostationary methods have been applied to such estimation, the development of *optimal* cyclostationary algorithms has been hindered in the same manner that optimal cyclic TDOA estimation algorithms were hindered. Yet the linear subspace model for deriving cyclostationary algorithms is applicable here as well. In particular, it should be possible to develop not only maximum likelihood AoA estimators, but their respective Cramér–Rao bounds as well.

All told, one change in modeling a digital communications signal has led to a whole host of new algorithms—the least of which was a new class of signal presence detectors. Particular algorithms, such as the overwhitener portion of the cyclic TDOA estimator and the cyclic ratio detectors, would have been difficult to derive without this new model. Finally, all of the new algorithms developed from this one simple change demonstrate the full value of using classical statistical principles—even when the underlying probability density function is only approximate.

Appendix A. Common Pulse Shapes

Several pulse shapes are commonly used to examine and model Pulse Amplitude Modulation (PAM) systems. Three particular pulse shapes, the rectangle pulse, Nyquist pulse, and the cosine pulse, are discussed here. Each of these three pulse shapes were used at some time during this research, so their definition is important. Therefore this appendix defines each pulse, and then expresses both in time and frequency. Along the way, figures will be presented to compare the pulse shapes in time and frequency.

The first pulse shape of interest is the rectangle pulse, shown in Fig. 58 and defined by,

$$\psi_{\mathrm{R}}(t) \triangleq \begin{cases} \frac{1}{\sqrt{T_s}} & -\frac{T_s}{2} < t < \frac{T_s}{2} \\ 0 & \text{otherwise} \end{cases} \tag{193}$$

This is perhaps the most common pulse for studying PAM waveforms. It is simple, and easy to work with theoretically [58]. Its properties, in terms of spectral redundancy, have been well studied and documented [21].

Unfortunately this pulse is not band–limited in frequency. This is easily seen by examining its Fourier transform,

$$\Psi_{\mathrm{R}}(f) = \sqrt{T_s}\,\mathrm{sinc}\,(fT_s). \tag{194}$$

Therefore, in order to use this pulse in practical applications, some amount of filtering is typically applied to this pulse prior to transmission. An ideal filter, applied to this pulse, would cut off all of the tails beyond $f = \pm\frac{1}{T_s}$. While this distorts the pulse shape in time, this may still be a good approximation to practice.

If being bandlimited is a requirement, the Nyquist pulse, also shown in Fig. 58, is a common solution. This pulse follows from the comments of Nyquist in [46], but is further developed in [63, 58]. In terms of this research, this pulse is defined in frequency as,

$$
\Psi_{\text{NYQ}}(f) \triangleq \begin{cases} \sqrt{T_s}\cos\left(\frac{\pi f T_s}{2}\right) & -\frac{1}{T_s} < f < \frac{1}{T_s} \\ 0 & \text{otherwise} \end{cases}
\tag{195}
$$

The fact that the Nyquist pulse is bandlimited in frequency carries with it the unfortunate side effect that it is infinite in time. This is easily seen from its time response below,

$$
\psi_{\text{NYQ}}(t) = \frac{1}{\sqrt{T_s}}\text{sinc}\left(\frac{2t}{T_s} - \frac{1}{2}\right) + \frac{1}{\sqrt{T_s}}\text{sinc}\left(\frac{2t}{T_s} + \frac{1}{2}\right).
\tag{196}
$$

This Nyquist pulse, as a function of time, is shown in Fig. 58. As you can see, though infinite, it tapers off to zero the farther you go from $t = 0$.

The truly unique and valuable feature of the Nyquist pulse is its time response after being convolved with itself. In particular, $\psi_{\text{NYQ}}(t) \otimes \psi_{\text{NYQ}}(t)$ has zero values at times of $t = kT_s$ for all integers k such that $k \neq 0$. That means that this bandlimited pulse shape will have zero ISI at the output of a matched filter, and no equalizer will be necessary [58].

A third type of pulse, having some of the benefits of both, is the cosine pulse, shown in Fig. 58 as well. This is the pulse used in PSK31 systems, and is known for the fact that alternating ones and zeros produce two pure tones [41]. It is simple to generate in time due to the short pulse length,

$$
\psi_{\cos}(t) = \begin{cases} \frac{1}{\sqrt{3T_s}} + \frac{1}{\sqrt{3T_s}}\cos\left(\frac{\pi t}{T_s}\right) & -T_s < t < T_s \\ 0 & \text{otherwise} \end{cases}
\tag{197}
$$
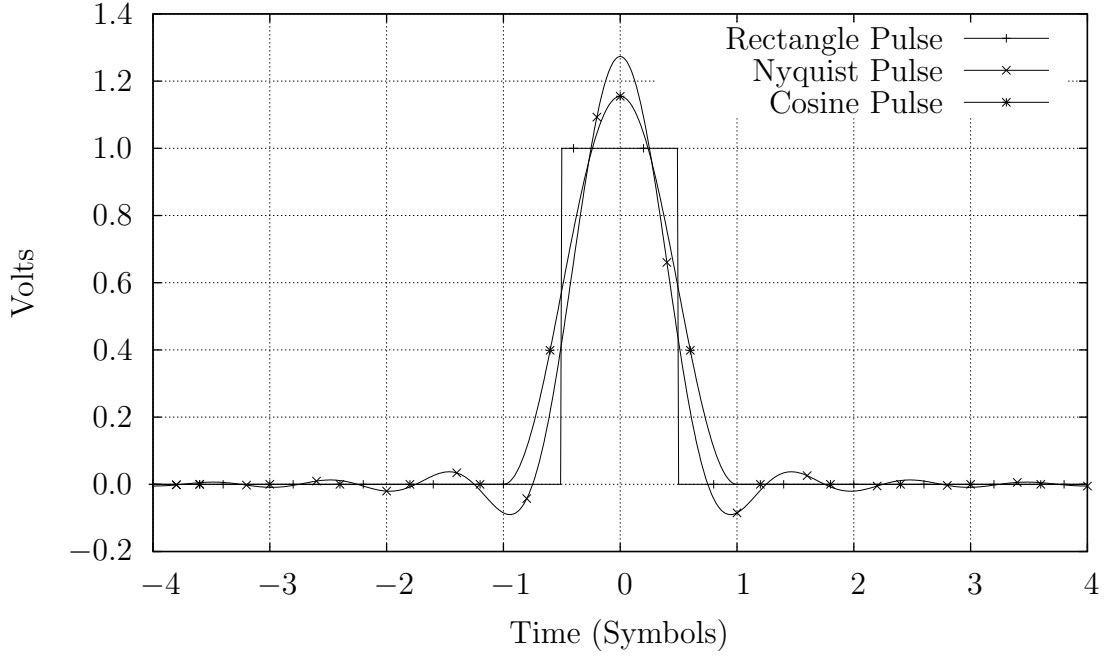
Figure 58.   Common Pulse Functions, in Time

Even better, this pulse tapers off quickly in frequency. The following demonstrates its frequency response.

$$\Psi_{\cos}(f) = 2\sqrt{\frac{T_s}{3}}\operatorname{sinc}(2fT_s) + \sqrt{\frac{T_s}{3}}\operatorname{sinc}[2fT_s - 1] + \sqrt{\frac{T_s}{3}}\operatorname{sinc}[2fT_s + 1]$$

(198)

This response is plotted in Fig. 59. As you can see from the figure, this pulse shape tapers quickly in frequency.

These three pulse functions are shown in the time domain in Fig. 58. Each pulse, derived above, has been normalized such that it has unit energy.

In frequency, these same pulse functions are shown in Fig. 59.

Following the formula given in Eqn. (7) on page 14, the PSD of signals generated using these pulse functions is proportional to the square of the pulse functions in frequency. These functions are plotted in Fig. 60. From Fig. 60, it is clearly seen
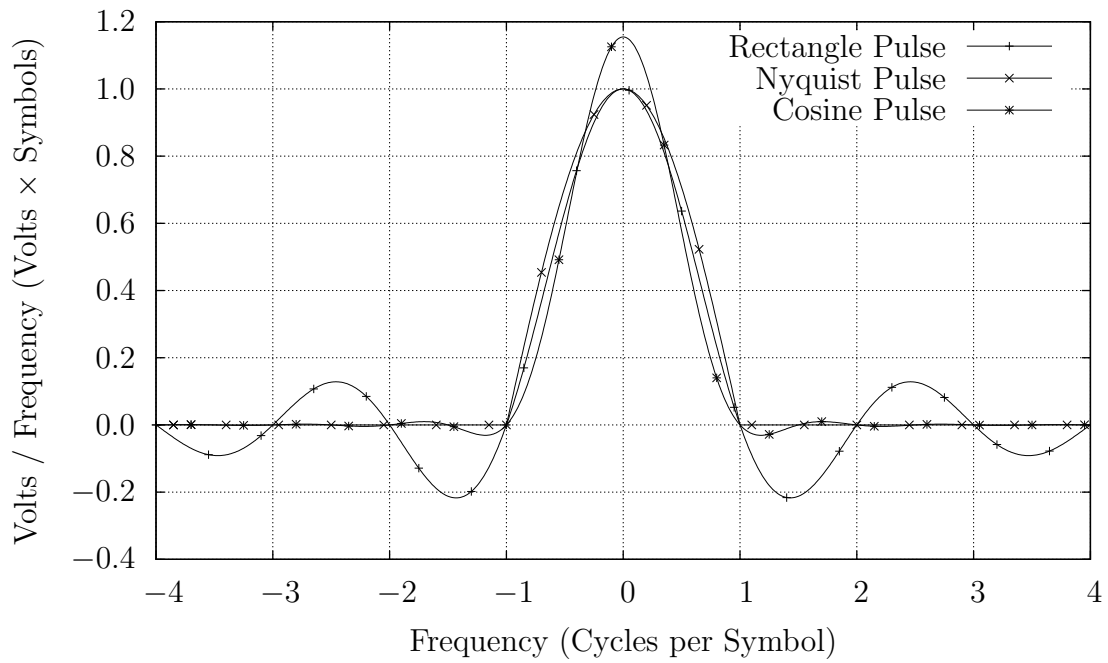
Figure 59.   Common Pulse Functions, in Frequency
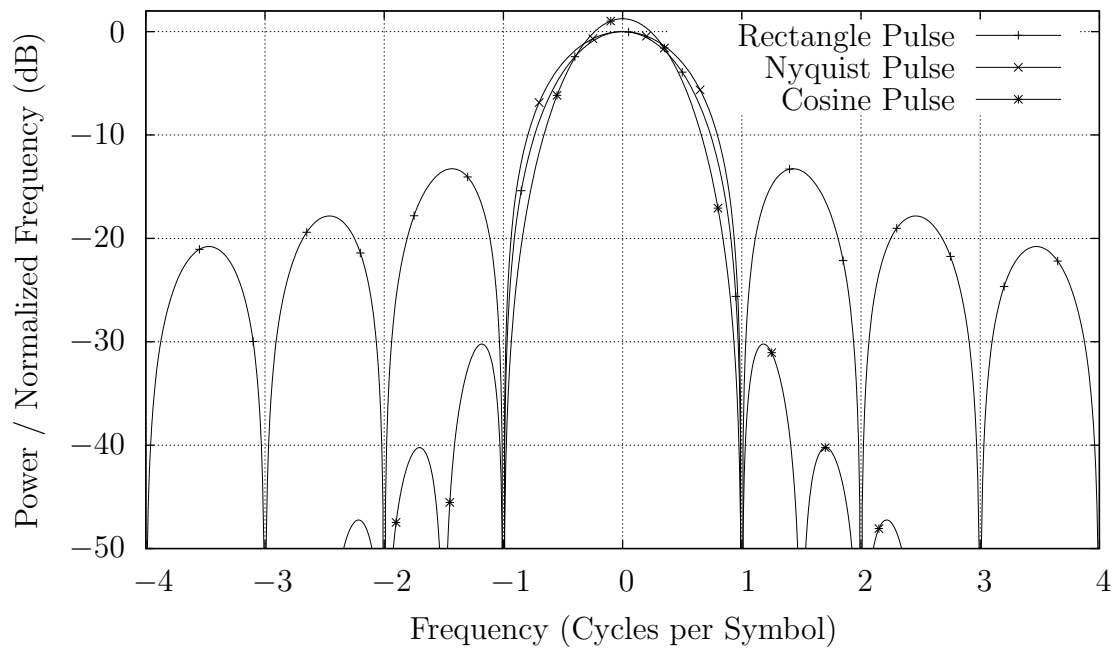
Figure 60.   Log PSD for Systems using Common Pulse Functions

189
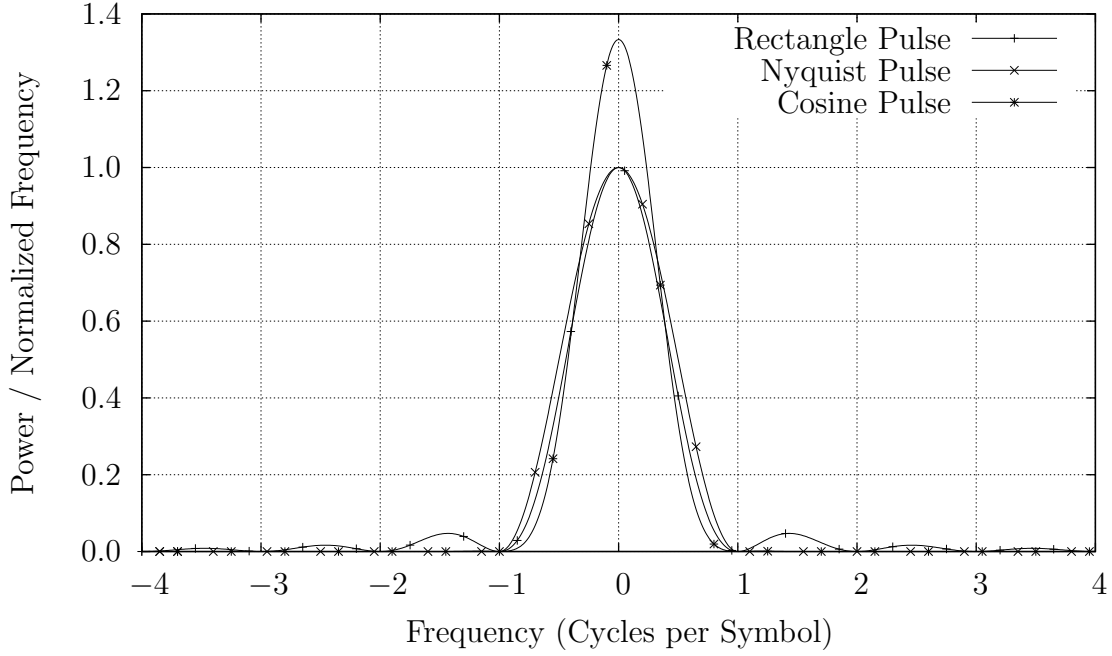
Figure 61. $\alpha = 0$ component of the Spectral Correlation Function

that the rectangle function does not taper off quickly as a function of frequency, while the cosine pulse tapers off much faster. The Nyquist pulse, on the other hand, tapers off to zero immediately. This demonstrates how a cosine pulse can be a cheap way of reducing sidelobes.

Since all of the algorithms within this dissertation depend in some manner upon the functions $|\Psi(f)|^2$ or $\Psi^*\left(f - \frac{1}{2T_s}\right)\Psi\left(f + \frac{1}{2T_s}\right)$, these two functions can be seen plotted in Figs. 61 and 62 respectively.

In reality, the actual choice of pulse shape is made by the design engineer. Nothing can be done about it in the receiver except compensating for it. When the pulse shape is known, compensation is simple and well defined. When the pulse function is unknown, it must either be guessed or estimated. Such estimation is well beyond the scope of this research.

190
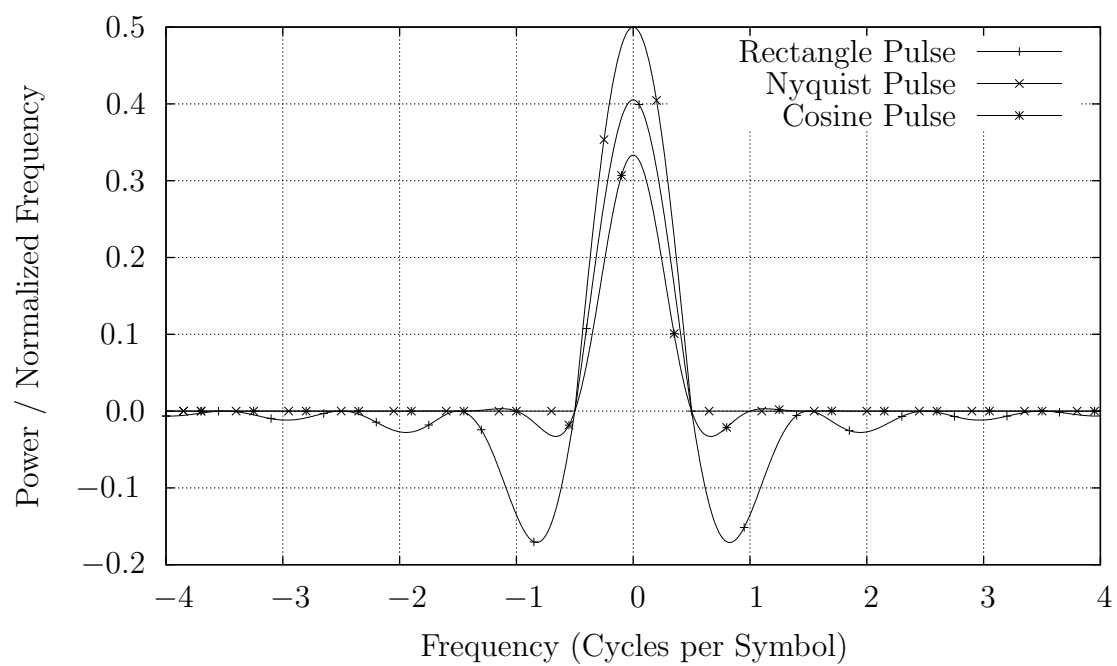
Figure 62. $\alpha = \frac{1}{T_s}$ component of the Spectral Correlation Function

Appendix B. Matrix Samples

Several matrices were introduced in Chap. III. However, it may not be obvious to the reader what the exact form of each of these matrices is. Therefore, this appendix presents samples of $\mathbf{R}_n$, $\mathbf{\Psi}$, and $\mathbf{R}_\phi$, such as would be used in the BPSK signaling case. From these examples, it should become obvious that $\mathbf{R}_n$ and $\mathbf{R}_\phi$ are both diagonal, and that $\mathbf{R}_\phi$ is unitary as well.

The first matrix sample is $\mathbf{R}_n$, a $4N_f \times 4N_f$ diagonal matrix representing the covariance of the noise,

$$
\mathbf{R}_n \approx T \begin{bmatrix}
S_n(f_1) & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots \\
0 & S_n(f_2) & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\
0 & 0 & \cdots & S_n(f_{N_f}) & 0 & \cdots & 0 & \cdots & 0 & \cdots \\
0 & 0 & \cdots & 0 & S_n\left(f_1 + \frac{1}{T_s}\right) & \cdots & 0 & \cdots & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\
0 & 0 & \cdots & 0 & 0 & \cdots & S_n(2f_c - f_1) & \cdots & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & S_n\left(2f_c - f_1 - \frac{1}{T_s}\right) & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots
\end{bmatrix}.
$$

$$\tag{199}$$

This matrix demonstrates, perhaps more than any other, the true reason for working in the frequency domain: it is diagonal. Had the problem been kept in the time domain, no such simplifying assumption could be made.

The next matrix, $\boldsymbol{\Psi}$, also a $4N_f \times N_f$ matrix, holds all of the pulse shaping information regarding the signal,

$$
\boldsymbol{\Psi} =
\begin{bmatrix}
\Psi\left(f_1 - f_c\right) & 0 & \cdots & 0 \\
0 & \Psi\left(f_2 - f_c\right) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \Psi\left(f_{N_f} - f_c\right) \\
\Psi\left(f_1 + \frac{1}{T_s} - f_c\right) & 0 & \cdots & 0 \\
0 & \Psi\left(f_2 + \frac{1}{T_s} - f_c\right) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \Psi\left(f_{N_f} + \frac{1}{T_s} - f_c\right) \\
\Psi^*\left(f_c - f_1\right) & 0 & \cdots & 0 \\
0 & \Psi^*\left(f_c - f_2\right) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \Psi^*\left(f_c - f_{N_f}\right) \\
\Psi^*\left(f_c - f_1 - \frac{1}{T_s}\right) & 0 & \cdots & 0 \\
0 & \Psi^*\left(f_c - f_2 - \frac{1}{T_s}\right) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \Psi^*\left(f_c - f_{N_f} - \frac{1}{T_s}\right)
\end{bmatrix} .
\tag{200}
$$

Example functions for $\Psi\left(f\right)$ are given in Appendix A. In general, $\Psi\left(f\right)$ is determined first by the signal of interest and second by the environment that it is placed within.

The final matrix, $\mathbf{R}_\phi$, also $4N_f \times 4N_f$ matrix, was used to contain all of the complex exponentials in the subspace equations,

$$\mathbf{R}_\phi \;=\; e^{j\theta}
\begin{bmatrix}
e^{-j2\pi(f_1-f_c)\tau} & 0 & \cdots & 0 & \cdots & \cdots & 0 & \cdots & 0 & \cdots \\
0 & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\
0 & \ddots & e^{-j2\pi\left(f_{N_f}-f_c\right)\tau} & 0 & \cdots & \cdots & 0 & \cdots & 0 & \cdots \\
0 & \cdots & 0 & e^{-j2\pi(f_1-f_c)\tau}e^{-j2\pi\frac{\tau}{T_s}} & 0 & \cdots & 0 & \cdots & 0 & \cdots \\
\vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots & \ddots & \vdots & \ddots \\
0 & \cdots & 0 & 0 & \cdots & \cdots & e^{-j2\theta}e^{-j2\pi(f_1-f_c)\tau} & \cdots & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\
0 & \cdots & 0 & 0 & \cdots & \cdots & 0 & \cdots & e^{-j2\theta}e^{-j2\pi(f_1-f_c)\tau}e^{-j2\pi\frac{\tau}{T_s}} & \cdots \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots
\end{bmatrix}
\tag{201}$$

In particular, these exponentials depend upon both τ and θ, values which need to be determined in order to specify $\mathbf{R}_\phi$. The utility of $\mathbf{R}_\phi$, even without knowing these values, is that it simplifies the linear equations. Further, the property that $\mathbf{R}_\phi^\dagger \mathbf{R}_\phi = \mathbf{I}$ is quite useful in simplifying algorithm development throughout this research.

As proved in Sec. 3.1, these matrices allow the algorithm designer to represent the signal as a subspace in frequency.

Appendix C. Redundancy Equations for an Offset QPSK System

The signal model studied in Sec. 3.1 assumed a Pulse Amplitude Modulated (PAM) signal. While this modulation type may be the most common, it is not the only digital modulation scheme. Other types of modulation exist that are not described by this model. This appendix proves that the subspace approach is robust enough to handle other modulation types as well, by demonstrating the extensions required to support Offset Quadrature Phase Shift Keying (OQPSK). Further, given that the difference between OQPSK and Minimum Shift Keying (MSK) lies in the choice of pulse function, $\psi(t)$, this section also extends this approach to MSK signaling as well. By developing the extension required for OQPSK, and then showing an example of the MMSE filter for an OQPSK system, the utility of this subspace approach to other modulation types should become more apparent.

As with Chapter III, this derivation starts with the definition of an OQPSK signal. In particular, an OQPSK can be represented as a sum of two PAM signals with the second one offset by half a symbol period and multiplied by an offset carrier [58],

$$
s(t) \;\; = \;\; A\Re \left\{ \begin{array}{l} \displaystyle\sum_{n=0}^{N_s-1} d_{in}\psi\left(t - nT_s - \tau\right) e^{j(2\pi f_c t + \theta)} \\[4pt] + \;\; j \displaystyle\sum_{n=0}^{N_s-1} d_{qn}\psi\left(t - nT_s - \frac{T_s}{2} - \tau\right) e^{j(2\pi f_c t + \theta)} \end{array} \right\} + n(t). \quad (202)
$$

The symbols have been split into two streams, d_{in} and d_{qn}, representing the in–phase and quadrature components respectively.

From this definition alone, an OQPSK signal can be viewed as a sum of two BPSK signals. The first BPSK signal is identical to the model used throughout this work, save that the symbols are now marked d_{in}. The second BPSK signal, however, also has a time delay, $\frac{T_s}{2}$, and a phase shift, $j = e^{j\frac{\pi}{2}}$, associated with it.

195

The Fourier transform of this signal can be evaluated as before as well. Except this time the Fourier transform is treated as the sum of two independent signals. The Fourier transform of this signal is,

$$
\begin{aligned}
s\left(f\right) \;=\; & \frac{e^{j\theta}}{2}e^{-j2\pi(f-f_c)\tau}\Psi\left(f-f_c\right)D_I\left(e^{j2\pi(f-f_c)T_s}\right) \\
& +j\frac{e^{j\theta}}{2}e^{-j2\pi(f-f_c)\tau}e^{-j2\pi(f-f_c)\frac{T_s}{2}}\Psi\left(f-f_c\right)D_Q\left(e^{j2\pi(f-f_c)T_s}\right)+n\left(t\right),
\end{aligned}
\tag{203}
$$

for positive frequencies, f. The two new terms in this equation, D_I and D_Q, correspond to the z–transforms of d_{in} and d_{qn} respectively. That is,

$$
D_I\left(e^{j2\pi(f-f_c)T_s}\right) \;\triangleq\; \sum_{n=0}^{N_s-1}d_{in}e^{-j2\pi(f-f_c)T_sn},
\tag{204}
$$

$$
\text{and } D_Q\left(e^{j2\pi(f-f_c)T_s}\right) \;\triangleq\; \sum_{n=0}^{N_s-1}d_{qn}e^{-j2\pi(f-f_c)T_sn}.
\tag{205}
$$

Using the fact that this signal is a sum of two independent BPSK signals, the redundancy equations can be stated as well. In particular,

$$
\begin{bmatrix}
X_s\left(f\right) \\
X_s\left(f+\frac{1}{T_s}\right) \\
X_s^*\left(2f_c-f\right) \\
X_s^*\left(2f_c-f-\frac{1}{T_s}\right)
\end{bmatrix}
=\; c_1
\begin{bmatrix}
\Psi\left(f-f_c\right) \\
e^{-j2\pi\frac{\tau}{T_s}}\Psi\left(f+\frac{1}{T_s}-f_c\right) \\
e^{-j2\theta}\Psi\left(f-f_c\right) \\
e^{-j2\theta}e^{-j2\pi\frac{\tau}{T_s}}\Psi\left(f+\frac{1}{T_s}-f_c\right)
\end{bmatrix}
D_I\left(e^{-j2\pi(f-f_c)T_s}\right)
$$

$$
+\; c_2
\begin{bmatrix}
\Psi\left(f-f_c\right) \\
-e^{-j2\pi\frac{\tau}{T_s}}\Psi\left(f+\frac{1}{T_s}-f_c\right) \\
-e^{-j2\theta}\Psi\left(f-f_c\right) \\
e^{-j2\theta}e^{-j2\pi\frac{\tau}{T_s}}\Psi\left(f+\frac{1}{T_s}-f_c\right)
\end{bmatrix}
D_Q\left(e^{-j2\pi(f-f_c)T_s}\right),
$$

$$
\tag{206}
$$

$$
\text{where } c_1 \;=\; \frac{Ae^{j\theta}e^{-j2\pi(f-f_c)\tau}}{2} \quad\text{and}\quad c_2 = jc_1e^{-j2\pi(f-f_c)\frac{T_s}{2}}.
$$

Further simplifying this equation can be done by defining two new quantities, $D_A\left(e^{j2\pi(f-f_c)T_s}\right)$ and $D_B\left(e^{j2\pi(f-f_c)T_s}\right)$, as

$$D_A\left(e^{j2\pi(f-f_c)T_s}\right) \triangleq D_I\left(e^{-j2\pi(f-f_c)T_s}\right) + je^{-j2\pi(f-f_c)\frac{T_s}{2}}D_Q\left(e^{-j2\pi(f-f_c)T_s}\right), \quad (207)$$

$$\text{and } D_B\left(e^{j2\pi(f-f_c)T_s}\right) \triangleq D_I\left(e^{-j2\pi(f-f_c)T_s}\right) - je^{-j2\pi(f-f_c)\frac{T_s}{2}}D_Q\left(e^{-j2\pi(f-f_c)T_s}\right). \quad (208)$$

Two sets of redundancy equations can then be created. The first set describes the left half of the signal's mainlobe,

$$\begin{bmatrix} X_s(f) \\ X_s^*\left(2f_c - f - \frac{1}{T_s}\right) \end{bmatrix} = \frac{Ae^{j\theta}e^{-j2\pi(f-f_c)\tau}}{2} \begin{bmatrix} \Psi(f-f_c) \\ e^{-j2\theta}e^{-j2\pi\frac{\tau}{T_s}}\Psi\left(f+\frac{1}{T_s}-f_c\right) \end{bmatrix} \\ \times D_A\left(e^{-j2\pi(f-f_c)T_s}\right),$$

$$(209)$$

while the second set describes the right half of the signal's mainlobe,

$$\begin{bmatrix} X_s\left(f+\frac{1}{T_s}\right) \\ X_s^*(2f_c - f) \end{bmatrix} = \frac{Ae^{j\theta}e^{-j2\pi(f-f_c)\tau}}{2} \begin{bmatrix} \Psi\left(f+\frac{1}{T_s}-f_c\right) \\ e^{-j2\theta}e^{-j2\pi\frac{\tau}{T_s}}\Psi(f-f_c) \end{bmatrix} \\ \times D_B\left(e^{-j2\pi(f-f_c)T_s}\right).$$

$$(210)$$

From this point, all of the analysis proceeds as before. The signal is rewritten as a matrix, and then the theory in Chapt. III applies.

In particular, the optimal filter for an OQPSK signal falls out in the exact same manner as before. This time, however, the form is slightly different. The optimal

197

filter for an OQPSK signal is,

$$H_{\text{MMSE}}(f) = \frac{\dfrac{A\Psi^*(f-f_c)}{2T_s S_n(f)}}{\dfrac{1}{2} + \dfrac{A^2}{4T_s}\dfrac{|\Psi(f-f_c)|^2}{S_n(f)} + \dfrac{A^2}{4T_s}\dfrac{\left|\Psi\left(f_c - f - \frac{1}{T_s}\right)\right|^2}{S_n\left(2f_c - f - \frac{1}{T_s}\right)}} \tag{211}$$

for $f \in \left(f_c - \frac{1}{T_s}, f_c\right)$ and

$$H_{\text{MMSE}}(f) = \frac{\dfrac{A\Psi^*(f-f_c)}{2T_s S_n(f)}}{\dfrac{1}{2} + \dfrac{A^2}{4T_s}\dfrac{|\Psi(f-f_c)|^2}{S_n(f)} + \dfrac{A^2}{4T_s}\dfrac{\left|\Psi\left(f_c - f + \frac{1}{T_s}\right)\right|^2}{S_n\left(2f_c - f + \frac{1}{T_s}\right)}} \tag{212}$$

for $f \in \left(f_c, f_c + \frac{1}{T_s}\right)$.

To demonstrate an example of this filter, the same noise and signal environment can be applied as in Sec. 4.2.1 and shown in Fig. 14 on page 116. Under these conditions, the optimal filter for an OQPSK signal is shown in Fig. 63. This is unlike the BPSK MMSE filter, where notching one noise frequency required boosting three other redundant frequencies. Instead, notching one frequency results in boosting the only other redundant frequency. The notch frequency and boost frequency are symmetric about either $f_c + \frac{1}{2T_s}$ or $f_c - \frac{1}{2T_s}$ depending upon which half of the band they are in.

This appendix has shown, therefore, that the theories presented in Chapt. III are easily extended to signaling types beyond PAM signals. The demonstration that this filter is indeed the optimal demodulation filter for an OQPSK signal has been rolled into further research.
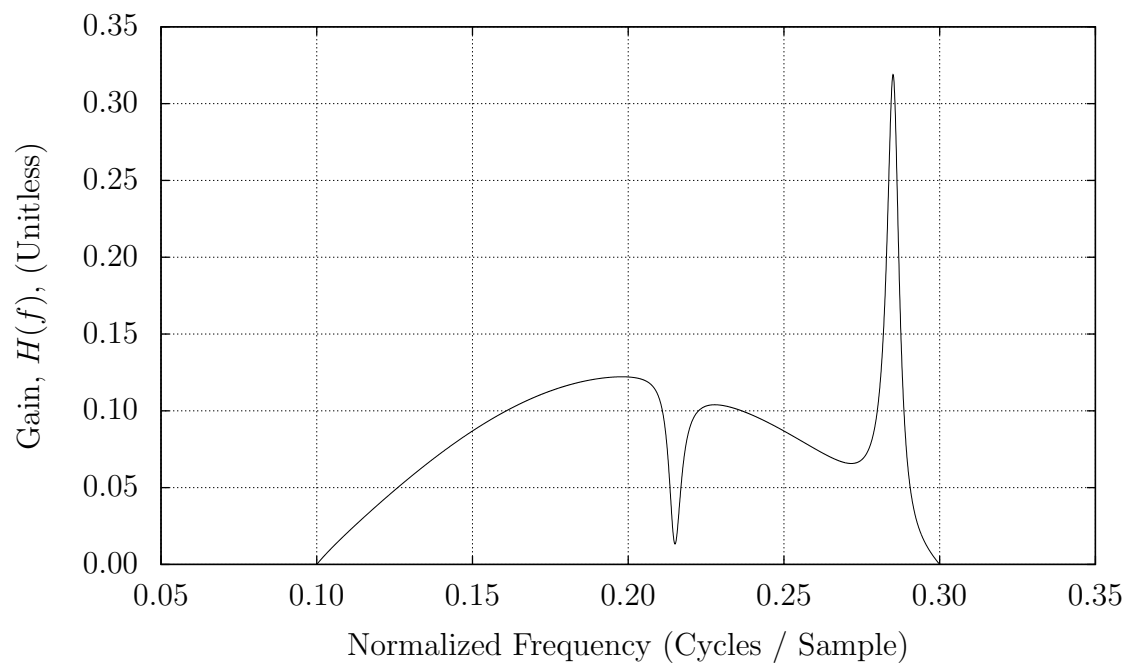
Figure 63.   MMSE Filter for an OQPSK Signal

199

Appendix D. Extensions to Binary Coherent Phase Signals

Laurent demonstrated that any binary coherent phase modulated signal (BCPM) can be expressed as a pulse amplitude modulated signal [39]. This expression has been used to recover the spectral correlation properties of such signals [44]. This section repeats his work, in order to derive expressions for the pulse functions of binary Coherent Phase Modulation (CPM) signals, in both time and frequency. Each of these expressions are then illustrated, in a second section, for Binary Frequency Shift Keyed (BFSK) systems.

D.1 General Binary Coherent Phase Modulation (BCPM)

BCPM signals are defined based upon some arbitrary time–limited function, $\phi(t)$, whose definition depends upon the signaling type. This function is defined over the interval $t \in [0, T_s)$ only. Given this function, the signal sent over any interval from kT_s to $(k+1)T_s$ is,

$$s(t) = \Re\left\{ e^{j\theta_k} \frac{e^{jp[k]\phi(t-kT_s)}}{\sqrt{T_s}} e^{j2\pi f_c t} \right\}, \tag{213}$$

where the $p[k] = \pm 1$ value is determined by which symbol is sent, and the $e^{j\theta_k}$ term is redefined to be the final phase of the previous symbol. This phase term is defined so that the phase of the total modulation remains continuous. This paper will derive the function, $\psi(t)$, and the values d_n, such that this same signal may be rewritten as,[1]

$$s(t) = \Re\left\{ \sum_n d_n \psi(t - nT_s) e^{j2\pi f_c t + j\theta} \right\}. \tag{214}$$

---

[1]τ has been dropped from this equation to make it easier to determine d_n, $\psi(t)$, and $\Psi(f)$. Once these have been determined, τ will need to be placed back into the model before applying any of the algorithms developed in Chap. III.

Further, the decomposition will be accomplished on the baseband signal,

$$s_{\mathrm{BB}}\left(t\right) \;=\; \frac{1}{\sqrt{T_s}}e^{j\theta}e^{jp[k]\phi(t-kT_s)}, \tag{215}$$

in order to simplify the analysis. Once accomplished, modulating the signal to bandpass frequencies may be accomplished as in Chapter III.

This decomposition will be accomplished via mathematical induction. Therefore, we start at the first symbol interval, from $0 \le t < T_s$,

$$s_{\mathrm{BB}}\left(t\right) \;=\; \frac{1}{\sqrt{T_s}}e^{j\theta_0}e^{jp[0]\phi(t)} \tag{216}$$

$$=\; \underbrace{\frac{1}{\sqrt{T_s}}e^{j\theta_0}\cos\left[\phi\left(t\right)\right]}_{\text{Data Independent}} + \underbrace{\frac{j}{\sqrt{T_s}}e^{j\theta_0}p\left[0\right]\sin\left[\phi\left(t\right)\right]}_{\text{Data Dependent}}. \tag{217}$$

Eqn. (217) shows that this symbol can be written as the sum of two components, a data independent component and a data dependent component. Likewise, the next symbol from $T_s \le t < 2T_s$ can also be written in terms of a data independent and a data dependent portion,

$$s_{\mathrm{BB}}\left(t\right) \;=\; \frac{1}{\sqrt{T_s}}e^{j\theta_1}\cos\left[\phi\left(t-T_s\right)\right] + \frac{j}{\sqrt{T_s}}e^{j\theta_1}p\left[1\right]\sin\left[\phi\left(t-T_s\right)\right]. \tag{218}$$

Substituting the value of $e^{j\theta_1}$,

$$e^{j\theta_1} \;=\; e^{j\theta_0}\cos\left[\phi\left(T_s\right)\right] + je^{j\theta_0}p\left[0\right]\sin\left[\phi\left(T_s\right)\right], \tag{219}$$

into this expression, this second symbol can be rewritten as,

$$s_{\text{BB}}(t) \ = \ \frac{1}{\sqrt{T_s}} e^{j\theta_0} \cos\left[\phi\left(T_s\right)\right] \cos\left[\phi\left(t - T_s\right)\right] \tag{220}$$

$$+ \ \frac{j}{\sqrt{T_s}} e^{j\theta_0} p\left[0\right] \sin\left[\phi\left(T_s\right)\right] \cos\left[\phi\left(t - T_s\right)\right]$$

$$+ \ \frac{j}{\sqrt{T_s}} e^{j\theta_1} p\left[1\right] \sin\left[\phi\left(t - T_s\right)\right].$$

Next, assume that the signal from $kT_s \leq t < (k+1) T_s$ can be written as,

$$s_{\text{BB}}(t) \ = \ \frac{1}{\sqrt{T_s}} e^{j\theta_0} \cos^k\left[\phi\left(T_s\right)\right] \cos\left[\phi\left(t - kT_s\right)\right] \tag{221}$$

$$+ \ \frac{j}{\sqrt{T_s}} \sin\left[\phi\left(T_s\right)\right] \cos\left[\phi\left(t - kT_s\right)\right] \sum_{m=0}^{k-1} e^{j\theta_m} p\left[m\right] \cos^{k-m-1}\left[\phi\left(T_s\right)\right]$$

$$+ \ \frac{j}{\sqrt{T_s}} e^{j\theta_k} p\left[k\right] \sin\left[\phi\left(t - kT_s\right)\right].$$

This is clearly true for $k = 1$, as seen in Eqn. 220. Next, we show that this is true for the interval from $(k + 1) T_s \leq t < (k + 2) T_s$. Over this interval,

$$
\begin{aligned}
s_{\mathrm{BB}}(t) \;=\;& s_{\mathrm{BB}}\left[(k + 1) T_s\right] e^{jp[k+1]\phi(t)} & (222) \\[2mm]
=\;& s_{\mathrm{BB}}\left[(k + 1) T_s\right]
\begin{pmatrix}
\cos\left[\phi\left(t - kT_s - T_s\right)\right] \\
+\quad jp\left[k + 1\right]\sin\left[\phi\left(t - kT_s - T_s\right)\right]
\end{pmatrix} \\[2mm]
=\;& \frac{1}{\sqrt{T_s}} e^{j\theta_0}\cos^{k+1}\left[\phi\left(T_s\right)\right]\cos\left[\phi\left(t - kT_s - T_s\right)\right] \\[2mm]
& +\quad \frac{j}{\sqrt{T_s}}\sin\left[\phi\left(T_s\right)\right]\cos\left[\phi\left(t - kT_s - T_s\right)\right]\sum_{m=0}^{k-1} e^{j\theta_m} p\left[m\right]\cos^{k-m}\left[\phi\left(T_s\right)\right] \\[2mm]
& +\quad \frac{j}{\sqrt{T_s}}\sin\left[\phi\left(T_s\right)\right]\cos\left[\phi\left(t - kT_s - T_s\right)\right]e^{j\theta_k} p\left[k\right] \\[2mm]
& +\quad \frac{j}{\sqrt{T_s}} e^{j\theta_{k+1}} p\left[k + 1\right]\sin\left[\phi\left(t - kT_s - T_s\right)\right] \\[2mm]
=\;& \frac{1}{\sqrt{T_s}} e^{j\theta_0}\cos^{k+1}\left[\phi\left(T_s\right)\right]\cos\left[\phi\left(t - kT_s - T_s\right)\right] & (223) \\[2mm]
& +\quad \frac{j}{\sqrt{T_s}}\sin\left[\phi\left(T_s\right)\right]\cos\left[\phi\left(t - kT_s - T_s\right)\right]\sum_{m=0}^{k} e^{j\theta_m} p\left[m\right]\cos^{k-m}\left[\phi\left(T_s\right)\right] \\[2mm]
& +\quad \frac{j}{\sqrt{T_s}} e^{j\theta_{k+1}} p\left[k + 1\right]\sin\left[\phi\left(t - kT_s - T_s\right)\right].
\end{aligned}
$$

Therefore, this relation holds for all time intervals, from $T_s \leq t$ to ∞.

Backing out the contribution from a particular value of $p[k]$ provides both the pulse function, $\psi(t)$, and the appropriate data symbol, d_n. In particular,

$$
d_n \;=\; js_{\mathrm{BB}}\left(nT_s\right) p\left[n\right], \text{ and} \tag{224}
$$

$$
\psi(t) \;=\;
\begin{cases}
0 & t < 0 \\[2mm]
\frac{1}{\sqrt{T_s}}\sin\left[\phi(t)\right] & 0 \leq t < T_s \\[2mm]
\frac{1}{\sqrt{T_s}}\sin\left[\phi\left(T_s\right)\right]\cos^{(k-1)}\left[\phi\left(T_s\right)\right]\cos\left[\phi\left(t - kT_s\right)\right] & kT_s \leq t < (k + 1) T_s
\end{cases}
\tag{225}
$$

All that remains is to take the Fourier transform of $\psi(t)$. This transform is immediately simplified by two definitions,

$$\Psi_{\mathrm{R}}(f) \triangleq \frac{1}{\sqrt{T_s}} \int_0^{T_s} \cos[\phi(t)] e^{-j2\pi ft} dt \tag{226}$$

$$\text{and } \Psi_{\mathrm{I}}(f) \triangleq \frac{1}{\sqrt{T_s}} \int_0^{T_s} \sin[\phi(t)] e^{-j2\pi ft} dt. \tag{227}$$

Having these two functions simplifies the Fourier transform of Eqn. (225) into a sum of components. In particular, the Fourier transform of the pulse function is,

$$\Psi(f) = \Psi_{\mathrm{I}}(f) + \Psi_{\mathrm{R}}(f) \sin[\phi(T_s)] \sum_{k=1}^{\infty} e^{-j2\pi fkT_s} \cos^{k-1}[\phi(T_s)] \tag{228}$$

$$= \Psi_{\mathrm{I}}(f) + \Psi_{\mathrm{R}}(f) \frac{\sin[\phi(T_s)] e^{-j2\pi fT_s}}{1 - \cos[\phi(T_s)] e^{-j2\pi fT_s}}. \tag{229}$$

This Fourier transform is vital to implementing the algorithms in this dissertation on any BCPM signal. First, from this pulse function, it is possible to define and implement the optimal MMSE filter. Further, this pulse function in frequency determines the form of the weights inside detectors and TDOA estimators. Neither of these are possible without this pulse function. Two correlations, in particular, are important. These are the zero–cycle correlation, or PSD, given by [21],

$$S_s(f) = \frac{A^2}{4T_s} |\Psi(f - f_c)|^2, \tag{230}$$

and the baud rate correlation, given by [21],

$$S_s^{\frac{1}{T_s}}(f) = e^{-j2\pi f \frac{\tau}{T_s}} \frac{A^2}{4T_s} \Psi^*\left(f - \frac{1}{2T_s} - f_c\right) \Psi\left(f - \frac{1}{2T_s} - f_c\right). \tag{231}$$

Examples of each of these pulse functions will be provided in the next section for BFSK systems.

Finally, it should be noted that this PAM decomposition has been demonstrated by others for arbitrary CPM modulation schemes—not just binary [42].

D.2 Binary Frequency Shift Keyed Systems

One important subclass within Binary Coherent Phase Systems is that of Binary Frequency Shift Keyed (BFSK) systems. Because this is an important subclass, this section accomplishes the PAM decomposition above for BFSK systems. Figures, presented throughout, will illustrate the shape of the pulse function, as well as its power and cross power spectral densities in frequency.

The first step to any CPM analysis, however, is to define the phase function, $\phi(t)$. In the case of a binary frequency shift keyed system, the system is either sending a tone at $f_c + \frac{h}{2T_s}$ or at $f_c - \frac{h}{2T_s}$. That means the phase function is linear,

$$\phi_{\text{FSK}}(t) = \frac{\pi h}{T_s}t. \tag{232}$$

The parameter, h, is called the *modulation index.*

Using the formula given in Eqn. (225), it is possible to plot these pulse functions, as in Fig. 64, for different values of the modulation index, h. This figure shows that, for $h = 0.5$, the pulse function extends over a full two symbols. Larger modulation indexes, such as $h = 0.6$ and $h = 0.7$, only cause this pulse function to overlap more symbols.
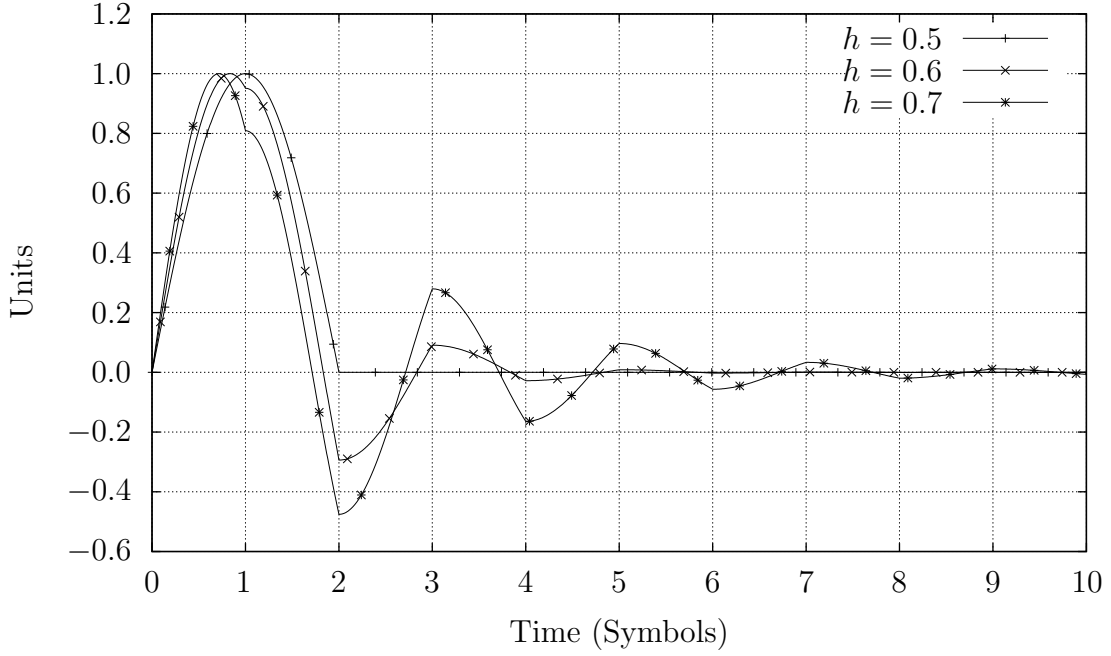
Figure 64.   BFSK Pulse Functions

In order to examine this pulse function in frequency, the functions $\Psi_I(f)$ and $\Psi_R(f)$ must first be determined. For BFSK systems, $\Psi_I(f)$ and $\Psi_R(f)$ are,

$$
\begin{aligned}
\Psi_I(f) &= - j\frac{\sqrt{T_s}}{2}e^{-j\pi(2fT_s - h)}\mathrm{sinc}\left[fT_s - \frac{h}{2}\right] \\
&+ j\frac{\sqrt{T_s}}{2}e^{-j\pi(2fT_s + h)}\mathrm{sinc}\left[fT_s + \frac{h}{2}\right],
\end{aligned}
\tag{233}
$$

and

$$
\begin{aligned}
\Psi_R(f) &= \frac{\sqrt{T_s}}{2}e^{-j\pi(2fT_s - h)}\mathrm{sinc}\left[fT_s - \frac{h}{2}\right] \\
&+ \frac{\sqrt{T_s}}{2}e^{-j\pi(2fT_s + h)}\mathrm{sinc}\left[fT_s + \frac{h}{2}\right].
\end{aligned}
\tag{234}
$$

Once Ψ_I and Ψ_R are known, it becomes a simple matter to calculate the spectral correlation functions of a BFSK signal.

The first spectral correlation function, the PSD, is shown in Fig. 65. This figure
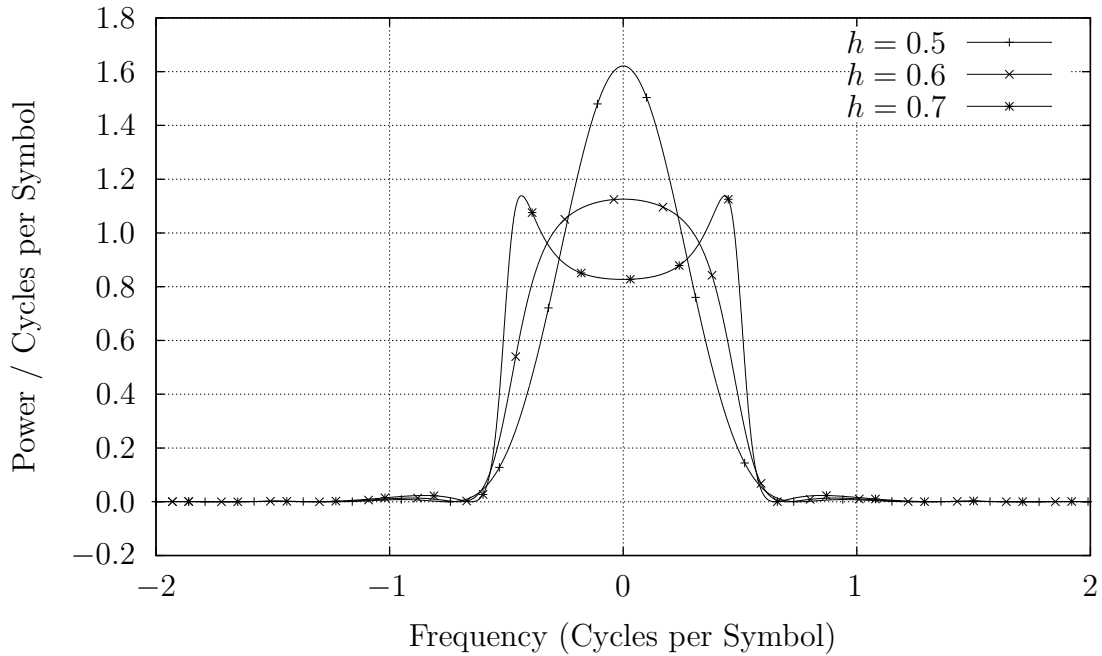
Figure 65.   BFSK Power Spectral Density

shows that as the modulation index increases, the signal goes from being focused in the center to being focused near $f_c \pm \frac{1}{2T_s}$. Looking at this same plot in decibels, Fig. 66, shows that as the signal widens the null–to–null bandwidth, $(2 - h)\frac{1}{T_s}$, decreases while the sidelobes increase as well.

The second spectral correlation function, the baud rate correlation, is shown in Fig. 67. Some important features need to be noted from this function. First, the zero crossings near the center occur at $\pm(1 - h)$. Second, as the modulation index increases, the edges of the signal become more and more highly correlated. At the same time the edges becomes more correlated, however, the sidelobes increase as well. Third, unlike the common pulse for phase shift keyed systems, the spectral correlation function does not go through zero at $f_c \pm \frac{1}{2T_s}$ (see Fig. 62). Practically, this makes it more difficult to arbitrarily declare an FSK signal to be bandlimited.

This appendix has presented Laurent's decomposition for binary CPM signals. This presentation derived the pulse functions, in time and frequency, that can be used
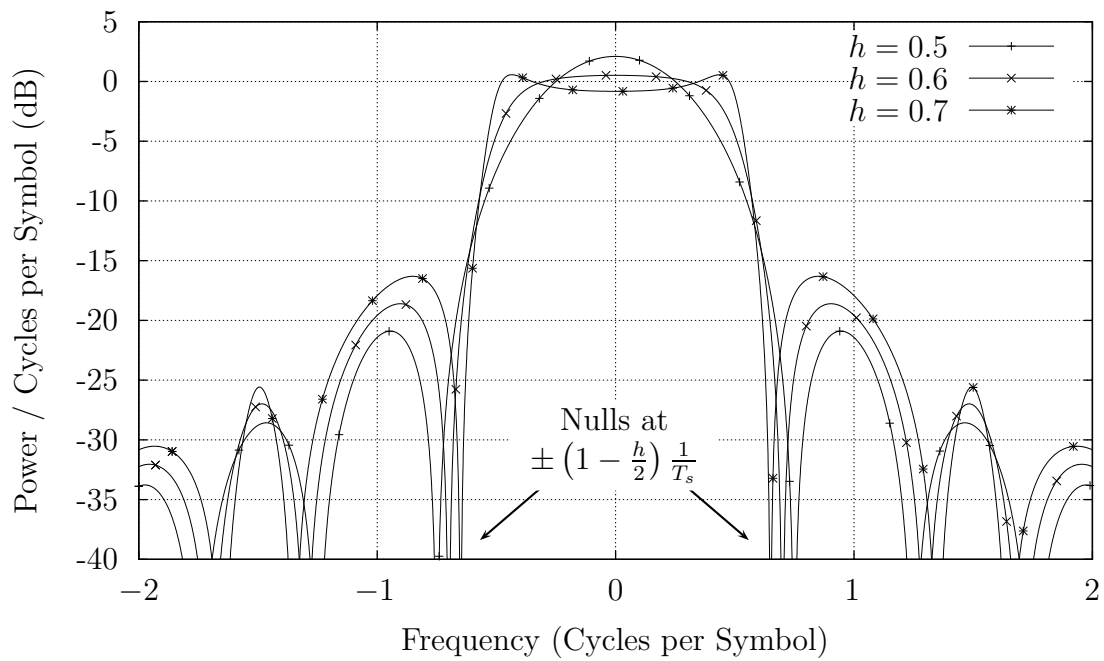
Figure 66. BFSK Power Spectral Density in dB
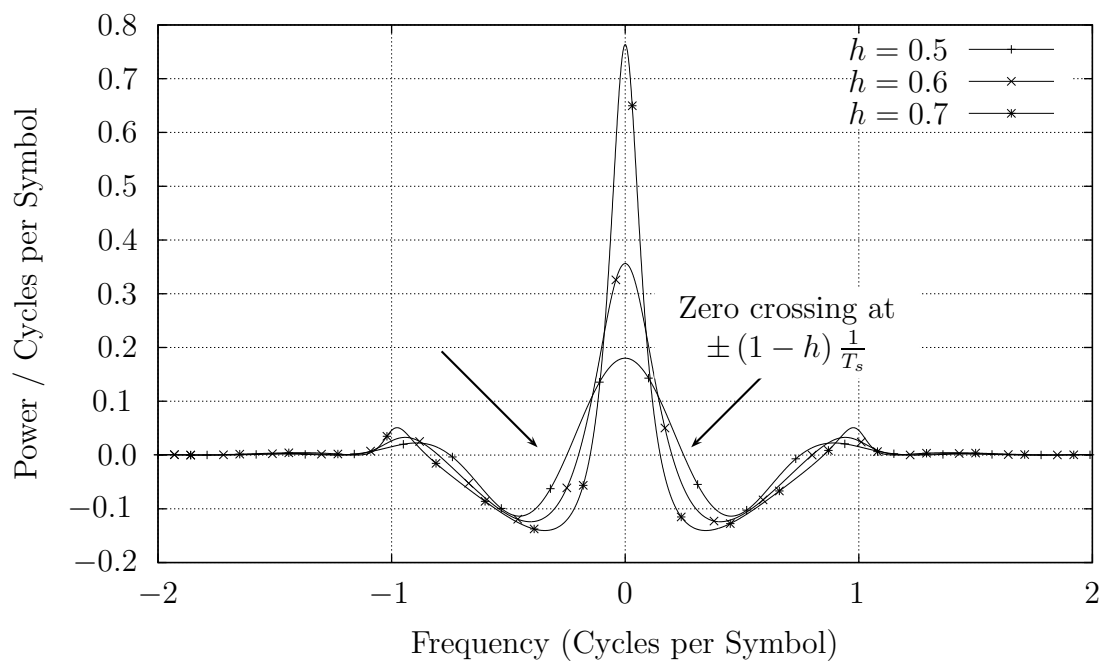
Figure 67. BFSK Spectral Correlation Function

208

with this description, together with the respective data symbols. These functions were then applied to BFSK systems to illustrate how they could be used to derive spectral correlations.

This representation connects BCPM signals with the rest of this research. This means that the MMSE filters presented earlier, together with the cyclic detectors and maximum likelihood TDOA estimators, all apply for BCPM signals. No modifications are required to these techniques to extend them to BCPM signals.

Appendix E. Differentiating with Respect to a Complex Vector

The development of $\hat{\boldsymbol{d}}_{\mathrm{MLE}}$, presented on page 72, depended upon taking the gradient of a function with respect to the conjugate of a vector quantity. Such gradients are not truly defined because the underlying conjugation operator is non–analytic. The main text has taken the approach, presented in [64], that such gradients can be calculated under certain circumstances. Although the simplicity of this approach makes it desirable, the fact that conjugation is non–analytic leaves the approach questionable. Therefore this appendix is offered to answer the question, what is the true maximum likelihood estimate of $\boldsymbol{d}$? As this appendix will show, the maximum likelihood estimate of $\boldsymbol{d}$ can be found using real vectors, thus avoiding the whole issue of taking the derivative of a non–analytic function.

Let $\boldsymbol{x}$ and $\boldsymbol{d}$ be complex vectors, $\mathbf{R}_n$ and $\mathbf{R}_d$ be valid covariance matrices, and let $\mathbf{C}$ be an arbitrary, possibly low rank, matrix. Finding the maximum likelihood estimate of $\boldsymbol{d}$ is then equivalent to maximizing the expression,[1]

$$\mathcal{L} \;=\; -\left(\boldsymbol{x} - \mathbf{C}\boldsymbol{d}\right)^{\dagger}\mathbf{R}_n^{-1}\left(\boldsymbol{x} - \mathbf{C}\boldsymbol{d}\right) - \boldsymbol{d}^{\dagger}\mathbf{R}_d\boldsymbol{d}, \tag{235}$$

with respect to the unknown parameter, $\boldsymbol{d}$. Since $\mathbf{R}_n$ and $\mathbf{R}_d$ are valid covariance matrices, they must be conjugate symmetric, $\mathbf{R}_n^{\dagger} = \left(\mathbf{R}_n^{T}\right)^{*}$. That means that the sum, Eqn. (235), above, must be real. As a real function, its derivative with respect to any individual element in $\boldsymbol{d}$ is well defined.

Therefore, maximizing $\mathcal{L}$ over $\boldsymbol{d}$ is equivalent to maximizing over both its real and imaginary components. By defining $\boldsymbol{d}_r = \Re\left\{\boldsymbol{d}\right\}$ and $\boldsymbol{d}_i = \Im\left\{\boldsymbol{d}\right\}$ such that,

$$\boldsymbol{d} \;=\; \boldsymbol{d}_r + j\boldsymbol{d}_i, \tag{236}$$

---

[1] In Sec. 3.1, the expression involved $\frac{A}{2}\mathbf{R}_{\phi}\boldsymbol{\Psi}_m\boldsymbol{d}$ and not $\mathbf{C}$. The slight change of notation has been made here for simplicity only.

each can be maximized independently. This leaves $\boldsymbol{d}_r$ and $\boldsymbol{d}_i$ real vectors and allows writing conjugation explicitly.

Since $\boldsymbol{d}_r$ is a vector, vector gradient formulas are applicable. The following are defined for real vectors, $\boldsymbol{d}_r$ (or equivalently $\boldsymbol{d}_i$).

$$\nabla_{\boldsymbol{d}_r} \boldsymbol{d}_r^T \boldsymbol{x} = \boldsymbol{x} \tag{237}$$

$$\nabla_{\boldsymbol{d}_r} \boldsymbol{x}^\dagger \boldsymbol{d}_r = \boldsymbol{x}^* \tag{238}$$

$$\nabla_{\boldsymbol{d}_r} \boldsymbol{d}_r^T \mathbf{R}_n \boldsymbol{d}_r = 2\mathbf{R}_n \boldsymbol{d}_r \tag{239}$$

Using these formulas, the location where $\nabla_{\boldsymbol{d}_r} \mathcal{L} = \nabla_{\boldsymbol{d}_i} \mathcal{L} = \boldsymbol{0}$ can be determined, and the vector $\hat{\boldsymbol{d}}$ can be defined as the complex vector whose real and imaginary components are solutions to this equation.

To get there, the components of $\mathcal{L}$ will be examined by breaking up the conjugation and expanding terms. This results in the following expression,

$$\mathcal{L} = -\left(\boldsymbol{x} - \mathbf{C}\boldsymbol{d}_r - j\mathbf{C}\boldsymbol{d}_i\right)^\dagger \mathbf{R}_n^{-1} \left(\boldsymbol{x} - \mathbf{C}\boldsymbol{d}_r - j\mathbf{C}\boldsymbol{d}_i\right) - \left(\boldsymbol{d}_r + j\boldsymbol{d}_i\right)^\dagger \mathbf{R}_d^{-1} \left(\boldsymbol{d}_r + j\boldsymbol{d}_i\right) \tag{240}$$

$$
\begin{aligned}
= \quad &- \quad \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \\
&+ \quad \boldsymbol{d}_r^T \mathbf{C}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \qquad + \quad \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{C}\boldsymbol{d}_r \\
&- \quad j\boldsymbol{d}_i^T \mathbf{C}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \qquad + \quad j\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{C}\boldsymbol{d}_i \\
&+ \quad j\boldsymbol{d}_i^T \mathbf{C}^\dagger \mathbf{R}_n^{-1} \mathbf{C}\boldsymbol{d}_r \quad - \quad j\boldsymbol{d}_r^T \mathbf{C}^\dagger \mathbf{R}_n^{-1} \mathbf{C}\boldsymbol{d}_i \\
&- \quad \boldsymbol{d}_r^T \mathbf{C}^\dagger \mathbf{R}_n^{-1} \mathbf{C}\boldsymbol{d}_r \quad - \quad \boldsymbol{d}_i^T \mathbf{C}^\dagger \mathbf{R}_n^{-1} \mathbf{C}\boldsymbol{d}_i \\
&- \quad \boldsymbol{d}_r^T \mathbf{R}_d^{-1} \boldsymbol{d}_r \qquad - \quad \boldsymbol{d}_i^T \mathbf{R}_d^{-1} \boldsymbol{d}_i \\
&- \quad j\boldsymbol{d}_r^T \mathbf{R}_d^{-1} \boldsymbol{d}_i \qquad + \quad j\boldsymbol{d}_i^T \mathbf{R}_d^{-1} \boldsymbol{d}_r
\end{aligned} \tag{241}
$$

From these terms, the gradient of $\mathcal{L}$ with respect to $\boldsymbol{d}_r$ and $\boldsymbol{d}_i$ can be examined.

First, looking at the gradient of $\mathcal{L}$ with respect to $\boldsymbol{d}_r$, we get

$$
\begin{aligned}
\nabla_{\boldsymbol{d}_r}\mathcal{L} &= \mathbf{C}^\dagger\mathbf{R}_n^{-1}\boldsymbol{x} + \left(\mathbf{C}^\dagger\mathbf{R}_n^{-1}\boldsymbol{x}\right)^* + j\left(\mathbf{C}^\dagger\mathbf{R}_n^{-1}\mathbf{C}\boldsymbol{d}_i\right)^* - j\mathbf{C}^\dagger\mathbf{R}_n^{-1}\mathbf{C}\boldsymbol{d}_i \\
&\quad - j\mathbf{R}_d^{-1}\boldsymbol{d}_i + j\left(\mathbf{R}_d^{-1}\boldsymbol{d}_i\right)^* - 2\mathbf{C}^\dagger\mathbf{R}_n^{-1}\mathbf{C}\boldsymbol{d}_r - 2\mathbf{R}_d^{-1}\boldsymbol{d}_r.
\end{aligned}
\tag{242}
$$

Solving for $\nabla\boldsymbol{d}_r = \mathbf{0}$ yields,

$$
\nabla_{\boldsymbol{d}_r}\mathcal{L} = 2\Re\left\{\mathbf{C}^\dagger\mathbf{R}_n^{-1}\boldsymbol{x}\right\} + 2\Im\left\{\mathbf{C}^\dagger\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right\}\boldsymbol{d}_i - 2\left(\mathbf{C}^\dagger\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right)\boldsymbol{d}_r
\tag{243}
$$

$$
\hat{\boldsymbol{d}}_r = \left(\mathbf{C}^\dagger\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right)^{-1}\left[\Re\left\{\mathbf{C}^\dagger\mathbf{R}_n^{-1}\boldsymbol{x}\right\} + \Im\left\{\mathbf{C}^\dagger\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right\}\boldsymbol{d}_i\right].
\tag{244}
$$

One particular term in this expression can be simplified here. That is $\Im\left\{\mathbf{C}^\dagger\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right\}$ must be zero. Consider first, that $\mathbf{R}_d^{-1}$ is a diagonal covariance matrix (see Sec. 3.1). Thus it must by necessity be real. That leaves $\mathbf{C}^\dagger\mathbf{R}_n^{-1}\mathbf{C}$. This term is real as well, but proving it is slightly more difficult and involves examining first the single sensor case and then the multi–sensor case.

In the single sensor case, $\mathbf{C} = \frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}$, and $\mathbf{R}_n$ is diagonal. When these terms are expanded out, they form a sum of positive numbers along each diagonal element, depending upon the modulation used and the number of redundancies. For the BPSK signal of interest, these terms turn into the sum,

$$
\begin{aligned}
\left(\frac{A^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\mathbf{R}_\phi\boldsymbol{\Psi}\right)_{ii} &= \frac{A^2}{4}\frac{|\Psi(f_i - f_c)|^2}{S_n(f_i)} + \frac{A^2}{4}\frac{\left|\Psi\left(f_i + \frac{1}{T_s} - f_c\right)\right|^2}{S_n\left(f_i + \frac{1}{T_s}\right)} \\
&\quad + \frac{A^2}{4}\frac{|\Psi(f_c - f_i)|^2}{S_n(2f_c - f_i)} + \frac{A^2}{4}\frac{\left|\Psi\left(f_c - f_i - \frac{1}{T_s}\right)\right|^2}{S_n\left(2f_c - f_i - \frac{1}{T_s}\right)},
\end{aligned}
\tag{245}
$$

which is clearly real.

The multi–sensor case, however, is a little bit more complicated. To show that $\mathbf{C}^\dagger \mathbf{R}_n^{-1} \mathbf{C}$ is a real diagonal matrix requires looking at $S_{n_i n_j}^\alpha (f)$. This function is related to the cross spectral density between the signals n_i and n_j. In particular,

$$S_{n_i n_j}^\alpha (f) \triangleq \lim_{T \to \infty} \frac{1}{T} \mathcal{E} \left\{ N_i^* \left(f - \frac{\alpha}{2} \right)_T N_j \left(f + \frac{\alpha}{2} \right)_T \right\}, \tag{246}$$

where $N_i(f)$ and $N_j(f)$ are the time–limited Fourier transforms of the noise plus interference on the i and j sensors respectively. When these terms are strictly stationary, regardless of any correlations between the sensors, then this function is strictly zero for all $\alpha \neq 0$. Going one step further, any linear combination of the sensors i and j will be uncorrelated for all frequency pairs but zero. Stated in equation form, for arbitrary constants c_i and c_j and $\alpha \neq 0$,

$$\lim_{T \to \infty} \mathcal{E} \left\{ \left[c_i N_i \left(f - \frac{\alpha}{2} \right)_T + c_j N_j \left(f - \frac{\alpha}{2} \right)_T \right]^* \left[c_i N_i \left(f + \frac{\alpha}{2} \right)_T + c_j N_j \left(f + \frac{\alpha}{2} \right)_T \right] \right\}$$
$$= c_i^2 S_{n_i}^\alpha (f) + c_j^2 S_{n_j}^\alpha (f) + c_i c_j S_{n_i n_j}^\alpha (f) + c_i c_j S_{n_j n_i}^\alpha (f)$$
$$= 0. \tag{247}$$

Carrying this back to $\mathbf{C}^\dagger \mathbf{R}_n^{-1} \mathbf{C}$, this means that prior to combining the redundant frequency components, the middle terms of this matrix are diagonal as well. Thus, in all stationary multi–sensor cases, $\mathbf{C}^\dagger \mathbf{R}_n^{-1} \mathbf{C}$ will refer to a diagonal covariance matrix. Since diagonal covariance matrices are real, this function must be real as well.

Now, using the fact that $\Im \left\{ \mathbf{C}^\dagger \mathbf{R}_n^{-1} \mathbf{C} \right\}$ is zero, the maximum likelihood estimate of $\boldsymbol{d}_r$ is,

$$\hat{\boldsymbol{d}}_r = \left(\mathbf{C}^\dagger \mathbf{R}_n^{-1} \mathbf{C} + \mathbf{R}_d^{-1} \right)^{-1} \Re \left\{ \mathbf{C}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \right\}. \tag{248}$$

The process to maximize $\mathcal{L}$ is repeated here for $\boldsymbol{d}_i$, and

$$
\begin{aligned}
\nabla_{\boldsymbol{d}_i}\mathcal{L} &= -j\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{x} + j\left(\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{x}\right)^* + j\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C}\boldsymbol{d}_r - j\left(\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C}\boldsymbol{d}_r\right)^* \\
&\quad + j\mathbf{R}_d^{-1}\boldsymbol{d}_r - j\left(\mathbf{R}_d^{-1}\right)^*\boldsymbol{d}_i - 2\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C}\boldsymbol{d}_i - 2\mathbf{R}_d^{-1}\boldsymbol{d}_i \\
&= 2\Im\left\{\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{x}\right\} - 2\Im\left\{\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right\}\boldsymbol{d}_r - 2\left(\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right)\boldsymbol{d}_i.
\end{aligned}
$$

$$(249)$$

Solving for $\nabla\boldsymbol{d}_i = \boldsymbol{0}$ yields,

$$
\begin{aligned}
\hat{\boldsymbol{d}}_i &= \left(\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right)^{-1}\left[\Im\left\{\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{x}\right\} - \Im\left\{\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right\}\boldsymbol{d}_r\right], \\
&= \left(\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right)^{-1}\Im\left\{\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{x}\right\}.
\end{aligned}
$$

$$(250)$$

Adding these estimates together yields the theoretical result,

$$
\hat{\boldsymbol{d}} = \left(\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C} + \mathbf{R}_d^{-1}\right)^{-1}\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\boldsymbol{x}.
$$

$$(251)$$

This result is identical to the one obtained in Sec. 3.2 by taking the gradient with respect to a complex vector quantity. This development, however, pulls out some assumptions that need to be made about the signal and noise covariances in order for this development to be valid. The first is that the signal covariance, $\mathbf{R}_d$, is diagonal. This is equivalent to saying that the d_n's form a discrete stationary sequence. The second is that the matrix, $\mathbf{C}^{\dagger}\mathbf{R}_n^{-1}\mathbf{C}$, is also diagonal. This is also equivalent to saying that linear combinations of the noise and interference made from any of the input sensors also form a stationary sequence. Both assumptions reduce to stationarity. If the symbol estimates, after sampling and downconversion, are stationary then this estimate holds.

Appendix F.   Theoretical verses Measured Detector Performance

Several approximations were presented in Sec. 2.4.1 in order to predict the performance of a radiometer in a white noise environment. This appendix revisits the particular approximation given in Eqn. (34), demonstrating that the radiometer detector performance shown earlier, in Fig. 49, at least comes close to matching its theoretical prediction.

The first step to calculating the theoretical performance of a radiometer using Eqn. (34), is to provide definitions for the values within that equation. These parameters, found in Table 8 on page 164, are re–iterated here for completeness. First, the burst length, $T = N_s T_s$, is 4096 samples. Second, the null–to–null bandwidth, $W = 2\frac{1}{T_s}$, is simply 0.125 cycles per sample. Together these create a time–bandwidth product, TW, of 512 (no units). Finally, recognizing that the ratio, $\frac{E}{N_o TW}$, in Eqn. (34) is simply another expression for the SNR in the burst, the probability of detection is then simply given by,

$$P_{\mathrm{D}} \quad \approx \quad Q\left[Q^{-1}\left(P_{\mathrm{FA}}\right) - \sqrt{TW}\mathrm{SNR}\right].$$

(252)

Given that this theoretical performance only becomes valid according to the central limit theorem as the time–bandwidth product, TW, increases towards infinity, there is no reason to expect a simulated radiometer with a smaller time–bandwidth product to exactly match this performance.

Now that all of the parameters in Eqn. (252) are known, all that remains is to compare this prediction to those shown earlier. Fig. 68 on the next page tells this story. This figure shows the performance of both the radiometer and the multicycle detector from the white noise simulation presented in Sec. 4.4, together with the predicted radiometer performance from Eqn. (252). As the theoretical prediction is only valid for large time–bandwidth products, Fig. 68 demonstrates a modest match

1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9

2
4
6
8
10
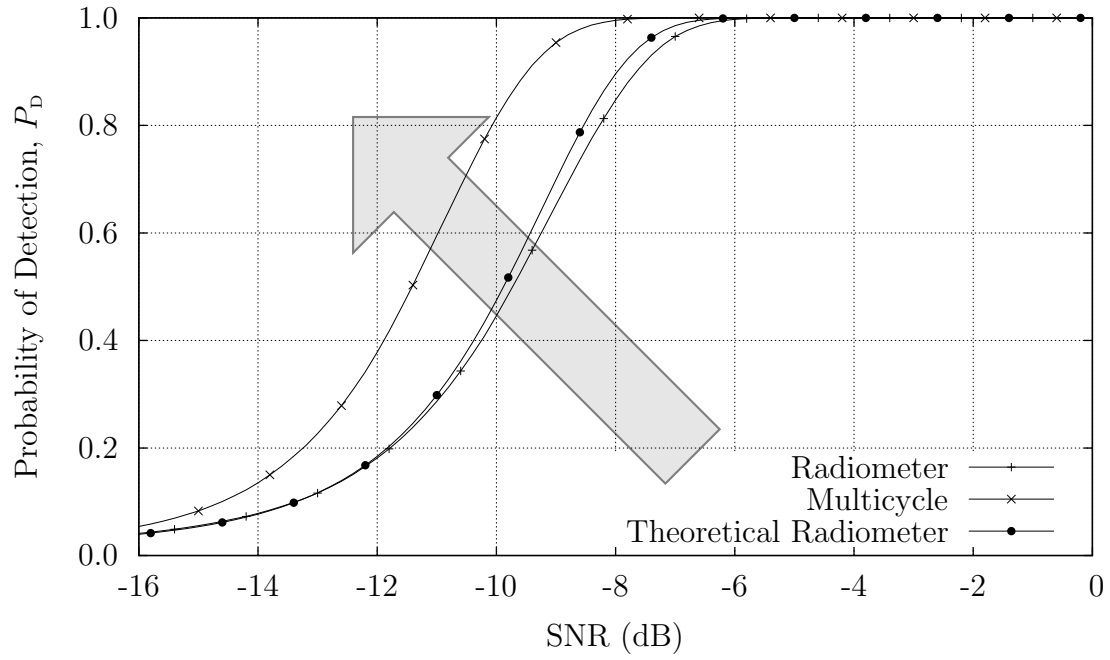
Figure 68.   Comparing simulated performance to theory

for a smaller time–bandwidth product. This commends the validity of the tests
presented in Sec. 4.4.

Appendix G. Deriving the Cyclic Ratio Detector

This appendix is a set of notes being used to derive a cyclic ratio detector. As with the other appendixes, the intention here is to be complete. Thus while the main documents skips steps for space, *every* step will be presented here.

Deriving the ratio detector requires starting at the likelihood function,

$$
\mathcal{L} \triangleq \ln f_{H_1}(\boldsymbol{x}, \boldsymbol{d}) \tag{253}
$$

$$
= -\frac{mN_f + N_f}{2} \ln(2\pi) - \frac{mN_f}{2} \ln \sigma^2 - \frac{1}{2} \ln \det |\mathbf{R}_n| - \frac{1}{2} \ln \det |\mathbf{R}_d| \tag{254}
$$

$$
-\frac{1}{2\sigma^2} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right)^\dagger \mathbf{R}_n^{-1} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right) - \frac{1}{2} \boldsymbol{d}^\dagger \mathbf{R}_d^{-1} \boldsymbol{d}
$$

and first deriving maximum likelihood estimates for $\boldsymbol{d}$,

$$
\nabla_{\boldsymbol{d}^\dagger} \mathcal{L} = \frac{A}{2\sigma^2} \mathbf{R}_\phi^\dagger \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} - \frac{A^2}{4\sigma^2} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} - \mathbf{R}_d^{-1} \boldsymbol{d} \tag{255}
$$

$$
= \frac{A}{2\sigma^2} \mathbf{R}_\phi^\dagger \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} - \left(\frac{A^2}{4\sigma^2} \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right) \boldsymbol{d} \tag{256}
$$

$$
\hat{\boldsymbol{d}}_{\mathrm{MLE}} = \frac{A}{2\sigma^2} \left(\frac{A^2}{4\sigma^2} \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \tag{257}
$$

and σ^2,

$$
\frac{\partial}{\partial \sigma^2} \mathcal{L} = -\frac{mN_f}{2\sigma^2} + \frac{1}{2\sigma^4} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right)^\dagger \mathbf{R}_n^{-1} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right) \tag{258}
$$

$$
\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{mN_f} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right)^\dagger \mathbf{R}_n^{-1} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right) \tag{259}
$$

$$
= \frac{1}{mN_f} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d}_{\mathrm{MLE}} \right)^\dagger \mathbf{R}_n^{-1} \left(\boldsymbol{x} - \frac{A}{2} \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d}_{\mathrm{MLE}} \right) \tag{260}
$$

If you simplify this expression for σ^2, by plugging in the ML estimate for $\boldsymbol{d}$, the following is the resulting estimator,

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{mN_f} \left[\begin{array}{l} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \\ - \ 2\frac{A^2}{4\sigma^2} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(\frac{A^2}{4\sigma^2} \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \\ + \ \frac{A^4}{16\sigma^4} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(\frac{A^2}{4\sigma^2} \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \\ \qquad \times \ \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} \left(\frac{A^2}{4\sigma^2} \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \end{array} \right]. \tag{261}$$

This estimate is still dependent upon σ^2, but such is unavoidable.

As in Sec. 3.4.2, define A_o to be $\frac{A}{2\sigma}$ to decouple the expressions for $\hat{\sigma}^2_{\text{MLE}}$ and $\hat{\boldsymbol{d}}_{\text{MLE}}$. Then,

$$\hat{\boldsymbol{d}}_{\text{MLE}} = \frac{A_o}{\sigma} \left(A_o^2 \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \mathbf{R}_\phi^\dagger \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}, \text{ and} \tag{262}$$

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{mN_f} \left[\begin{array}{l} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \\ - \ 2A_o^2 \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(A_o^2 \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \\ + \ A_o^4 \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(A_o^2 \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \\ \qquad \times \ \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} \left(A_o^2 \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \end{array} \right]. \tag{263}$$

Recall the definition of a locally most powerful detector. Such a detector is defined, for a generic parameter A near $A = 0$, by

$$y(\boldsymbol{x}) \triangleq \frac{\frac{\partial}{\partial A} f(\boldsymbol{x}|A)\big|_{A=0}}{f(\boldsymbol{x}|0)}. \tag{264}$$

Using $f(\boldsymbol{x}) = \exp\{\mathcal{L}\}$, this detection statistic is easily simplified in terms of $\mathcal{L}$,

$$\frac{\frac{\partial}{\partial A} f(\boldsymbol{x}|A)\big|_{A=0}}{f(\boldsymbol{x}|0)} = \frac{\partial}{\partial A} \mathcal{L}\bigg|_{A=0} \tag{265}$$

In this particular case, we would like to find the optimal detector near $A_o^2 = 0$. Plugging in values, we see that this detector only depends upon three terms from the original PDF. Of these three terms, the first term (below) is, by nature of the ML estimate from which it was derived, constant and not dependent upon A_o at all.

$$\frac{\partial}{\partial A_o^2} \mathcal{L} = \frac{\partial}{\partial A_o^2} \left[\begin{array}{c} -\frac{mN_f \hat{\sigma}_{\text{MLE}}^2}{2\hat{\sigma}_{\text{MLE}}^2} \\ -\frac{1}{2} \hat{\boldsymbol{d}}_{\text{MLE}}^\dagger \mathbf{R}_d^{-1} \hat{\boldsymbol{d}}_{\text{MLE}} \\ -\frac{mN_f}{2} \ln \hat{\sigma}_{\text{MLE}}^2 \end{array} \right]\Bigg|_{A_o=0} \tag{266}$$

The second two expand to yield the detector of interest,

$$\frac{\partial}{\partial A_o^2} \mathcal{L} = \frac{\partial}{\partial A_o^2} \left[\begin{array}{c} -\frac{A_o^2}{2\hat{\sigma}_{\text{MLE}}^2} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(A_o^2 \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)^{-1} \\ \times \mathbf{R}_d^{-1} \left(A_o^2 \boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x} \end{array} \right]\Bigg|_{A_o=0}$$
$$- \frac{mN_f}{2} \frac{\frac{\partial}{\partial A_o^2} \hat{\sigma}_{\text{MLE}}^2 \big|_{A_o=0}}{\frac{1}{mN_f} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}}$$
$$= -\frac{\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}}{2 \frac{1}{mN_f} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}} + \frac{mN_f}{2} \frac{\frac{2}{mN_f} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}}{\frac{1}{mN_f} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}} \tag{267}$$
$$= \frac{mN_f}{2} \frac{\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}}{\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}} \tag{268}$$

Removing the unnecessary scale constant from this detector results in the detector shown in Eqn. (176) on page 101,

$$y_{\text{RATIO}}(\boldsymbol{x}) = \frac{\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}}{\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}} \tag{269}$$

Appendix H. Handling Angles in BPSK Detectors

One particular problem makes implementing the BPSK multicycle detector difficult, and that is determining the angles τ and θ. This problem is known, and has been lamented before,

> Unfortunately, even if the modulation type and its parameter values (e.g., carrier frequency and chip rate for a BPSK signal) are known, the optimum multicycle detector cannot be implemented without knowledge of the phase of the signal because the quantities $S_s^\alpha(f)$ depend on this phase. [17, p. 903]

Yet in order to implement the multicycle detector presented in Sec. 4.4, these angles needed to be determined. This appendix details the method which was used in this research to determine those angles.

Starting from the definition of the multicycle detector, the detection statistic is,

$$
y_{\text{MCYC}}\left(\boldsymbol{x}\right) \;=\; \max_{\tau,\theta} \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}. \tag{270}
$$

This expression can be broken down into components, in order to separate and simplify the problem of determining τ and θ. Doing this results in an expression composed of five separate terms,

$$
y_{\text{MCYC}}\left(\boldsymbol{x}\right) \;=\; C_0 + \max_{\tau,\theta} 2\Re \left\{ \begin{array}{ccc} e^{j2\pi \frac{\tau}{T_s}} C_1 & + & e^{j2\theta} C_2 \\ + \; e^{j2\theta} e^{j2\pi \frac{\tau}{T_s}} C_3 & + & e^{j2\theta} e^{-j2\pi \frac{\tau}{T_s}} C_4 \end{array} \right\}, \tag{271}
$$

where the C_is are independent of τ and θ,

$$
\begin{aligned}
C_0 \;=\; &\sum_{f_i} \left|\Psi\left(f_i - f_c\right)\right|^2 \frac{|X(f_i)|^2}{S_n^2(f_i)} \\
&+\; \left|\Psi\left(f_i + \tfrac{1}{T_s} - f_c\right)\right|^2 \frac{\left|X\left(f_i + \tfrac{1}{T_s}\right)\right|^2}{S_n^2\left(f_i + \tfrac{1}{T_s}\right)} \\
&+\; \left|\Psi\left(f_i - f_c\right)\right|^2 \frac{|X(2f_c - f_i)|^2}{S_n^2(2f_c - f_i)} \\
&+\; \left|\Psi\left(f_i + \tfrac{1}{T_s} - f_c\right)\right|^2 \frac{\left|X\left(2f_c - f_i - \tfrac{1}{T_s}\right)\right|^2}{S_n^2\left(2f_c - f_i - \tfrac{1}{T_s}\right)}
\end{aligned}
\tag{272}
$$

$$
\begin{aligned}
C_1 \;=\; &\sum_{f_i} \Psi\left(f_i - f_c\right)\Psi^*\left(f_i + \tfrac{1}{T_s} - f_c\right) \frac{X^*(f_i)X\left(f_i + \tfrac{1}{T_s}\right)}{S_n(f_i)S_n\left(f_i + \tfrac{1}{T_s}\right)} \\
&+\; \Psi\left(f_i - f_c\right)\Psi^*\left(f_i + \tfrac{1}{T_s} - f_c\right) \frac{X(2f_c - f_i)X^*\left(2f_c - f_i - \tfrac{1}{T_s}\right)}{S_n(2f_c - f_i)S_n\left(2f_c - f_i - \tfrac{1}{T_s}\right)}
\end{aligned}
\tag{273}
$$

$$
\begin{aligned}
C_2 \;=\; &\sum_{f_i} \left|\Psi\left(f_i - f_c\right)\right|^2 \frac{X^*(f_i)X^*(2f_c - f_i)}{S_n(f_i)S_n(2f_c - f_i)} \\
&+\; \left|\Psi\left(f_i + \tfrac{1}{T_s} - f_c\right)\right|^2 \frac{X^*\left(f_i + \tfrac{1}{T_s}\right)X^*\left(2f_c - f_i - \tfrac{1}{T_s}\right)}{S_n\left(f_i + \tfrac{1}{T_s}\right)S_n\left(2f_c - f_i - \tfrac{1}{T_s}\right)}
\end{aligned}
\tag{274}
$$

$$
C_3 \;=\; \sum_{f_i} \Psi\left(f_i - f_c\right)\Psi^*\left(f_i + \frac{1}{T_s} - f_c\right) \frac{X^*(f_i)\,X^*\left(2f_c - f_i - \tfrac{1}{T_s}\right)}{S_n(f_i)\,S_n\left(2f_c - f_i - \tfrac{1}{T_s}\right)}
\tag{275}
$$

$$
\text{and } C_4 \;=\; \sum_{f_i} \Psi\left(f_i + \frac{1}{T_s} - f_c\right)\Psi^*\left(f_i - f_c\right) \frac{X^*\left(f_i + \tfrac{1}{T_s}\right)X^*(2f_c - f_i)}{S_n\left(f_i + \tfrac{1}{T_s}\right)S_n(2f_c - f_i)}.
\tag{276}
$$

While each of these terms can be used individually as a single cycle detector, using them together requires finding the maximum of Eqn. (271).

Factoring $e^{j2\theta}$ from the latter three terms in this expression,

$$
y_{\text{MCYC}}\left(\boldsymbol{x}\right) \;=\; C_0 + \max_{\tau,\theta} 2\Re\left\{e^{j2\pi\frac{\tau}{T_s}}C_1\right\} + 2\Re\left\{e^{j2\theta}\left(C_2 + e^{j2\pi\frac{\tau}{T_s}}C_3 + e^{-j2\pi\frac{\tau}{T_s}}C_4\right)\right\},
\tag{277}
$$

yields an equation that is easily maximized for θ. Thus,

$$y_{\text{MCYC}}(\boldsymbol{x}) \;=\; C_0 + \max_{\tau} 2\Re\left\{e^{j2\pi\frac{\tau}{T_s}}C_1\right\} + \left|C_2 + e^{j2\pi\frac{\tau}{T_s}}C_3 + e^{-j2\pi\frac{\tau}{T_s}}C_4\right|, \quad (278)$$

is the maximum value across all θ.

The problem of maximizing this expression over all τ remains. Since no analytical solution appears possible at this point, numerical methods must be used to solve this equation. One useful numerical method is Newton's method for root–finding [1, 48]. This method can be used to find the value of τ which makes,

$$\frac{\partial}{\partial\tau} y_{\text{MCYC}}(\boldsymbol{x}) \;=\; 0. \tag{279}$$

To use Newton's method, however, an initial choice needs to be made for τ. Calling this initial estimate $\hat{\tau}_0$, and recognizing that C_1 is going to be the largest term in this expression,

$$\hat{\tau}_0 \;\triangleq\; \frac{C_1^*}{|C_1|}, \tag{280}$$

appears to be a particularly good choice.

From here, the next step is to apply Newton's method to the derivative,

$$\frac{\partial}{\partial\tau} y_{\text{MCYC}}(\boldsymbol{x}) \;=\; -2\frac{2\pi}{T_s}\Im\left\{e^{j2\pi\frac{\tau}{T_s}}C_1\right\} - \frac{2\pi}{T_s}\frac{\Im\left\{e^{j2\pi\frac{\tau}{T_s}}\left[C_2^*C_3 + C_2 C_4^*\right] + 2e^{j4\pi\frac{\tau}{T_s}}C_4^*C_3\right\}}{\left|C_2 + e^{j2\pi\frac{\tau}{T_s}}C_3 + e^{-j2\pi\frac{\tau}{T_s}}C_4\right|},$$

$$\tag{281}$$

since it will be zero at the location of any maximum. To drive a solution towards this root, the second derivative with respect to τ is also required,

$$
\begin{aligned}
\frac{\partial^2}{\partial \tau^2} y_{\text{MCYC}}(\boldsymbol{x}) = & - 2 \left(\frac{2\pi}{T_s} \right)^2 \Re \left\{ e^{j2\pi \frac{\tau}{T_s}} C_1 \right\} \\
& - \left(\frac{2\pi}{T_s} \right)^2 \frac{\Re \left\{ e^{j2\pi \frac{\tau}{T_s}} [C_2^* C_3 + C_2 C_4^*] + 4 e^{j4\pi \frac{\tau}{T_s}} C_4^* C_3 \right\}}{\left| C_2 + e^{j2\pi \frac{\tau}{T_s}} C_3 + e^{-j2\pi \frac{\tau}{T_s}} C_4 \right|} \\
& - 3 \left(\frac{2\pi}{T_s} \right)^2 \frac{\Im \left\{ e^{j2\pi \frac{\tau}{T_s}} [C_2^* C_3 + C_2 C_4^*] + 2 e^{j4\pi \frac{\tau}{T_s}} C_4^* C_3 \right\}^2}{\left| C_2 + e^{j2\pi \frac{\tau}{T_s}} C_3 + e^{-j2\pi \frac{\tau}{T_s}} C_4 \right|^3} .
\end{aligned}
\tag{282}
$$

Then, according to Newton's method, the following expression should converge to the nearest root [1],

$$
\hat{\tau}_{n+1} = \hat{\tau}_n - \frac{\frac{\partial}{\partial \tau} y_{\text{MCYC}}(\boldsymbol{x})}{\frac{\partial^2}{\partial \tau^2} y_{\text{MCYC}}(\boldsymbol{x})} .
\tag{283}
$$

One problem remains, however, with this method: it doesn't work. While it does correctly find a root in $\frac{\partial}{\partial \tau} y_{\text{MCYC}}(\boldsymbol{x})$, it does not necessarily find the *maximum*. To make this solution converge to a maximum, one slight modification is made to Newton's formula,

$$
\hat{\tau}_{n+1} = \hat{\tau}_n + \frac{\frac{\partial}{\partial \tau} y_{\text{MCYC}}(\boldsymbol{x})}{\left| \frac{\partial^2}{\partial \tau^2} y_{\text{MCYC}}(\boldsymbol{x}) \right|} .
\tag{284}
$$

This expression always converges to a maximum value. While its convergence might not start out very fast, and indeed it does not, typically within ten iterations the expression for τ, and likewise for $y_{\text{MCYC}}(\boldsymbol{x})$, has converged enough to be usable.

Thus, from Eqn. (284), we have an expression for τ and from Eqn. (278), we have a maximum likelihood estimate for θ. Putting these two estimates into Eqn. (278) provides an estimate of $y_{\text{MCYC}}(\boldsymbol{x})$ which can be used in a detector. This

method for dealing with the nuisance parameters τ, and θ was used in the detection tests of Sec. 4.4. The two facts that the multicycle detector outperformed all other detectors when the signal was present, and that it rejected more interference than the radiometer when the signal was not present, both commend the capability and utility of this method.

Appendix I. Subspace Detection

All of the cyclostationary detectors, such as those presented in in Sec. 3.4.2 as well as those developed in [17], are *locally* most powerful detectors. That means that, when the signal is weak, no other detector has a greater probability of detection for a fixed probability of false alarm [13]. Yet as the signal strength increases, no such claim to optimality can be made. A more desirable detector would be a *uniformly* most powerful (UMP) detector—one that has a greater probability of detecting the signal than every other detector independent of the strength of the signal.

From linear subspace theory, Scharf and Friedlander have developed matched subspace detectors that are UMP regardless of the strength of the signal within the subspace [54]. These detectors are very similar in form to the detectors presented in Sec. 3.4. Further, they are provably the optimal detectors for detecting unknown signals lying within a known subspace—regardless of signal strength. This section presents these matched subspace detectors, and tests them in the signal environment to determine which detectors, whether the subspace detectors or the locally optimal cyclostationary detectors, are truly optimal.

The subspace detectors follow after changing two assumptions. The first is that the subspace containing the signal, including $\mathbf{R}_\phi$ and thus τ and θ, is fully known. The second changed assumption is that $\boldsymbol{d}$ is unknown and deterministic, as opposed to random vector with a known distribution. Under these assumptions, the hypothesis test is framed as,

$$H_0 \quad : \quad \boldsymbol{x} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{R}_n\right) \tag{285}$$

$$\text{versus } H_1 \quad : \quad \boldsymbol{x} \sim \mathcal{N}\left(\frac{A}{2}\mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d}, \sigma^2 \mathbf{R}_n\right), \ A > 0. \tag{286}$$

The most obvious difference between this hypothesis test and those in Eqns. (166) and (167), is that the data vector, $\boldsymbol{d}$, no longer has a distribution associated with it.

A more subtle difference is that the matrix $\mathbf{R}_\phi$ is fully specified in this test. From this hypothesis test, the UMP detector for determining if $A > 0$ is given by [54]

$$y_{\text{SUBS-KNOWN}}(\boldsymbol{x}) \;=\; \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(\boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}. \tag{287}$$

Using the same hypotheses, only this time letting σ^2 be unknown, a similar detector to y_{RATIO} results,

$$y_{\text{SUBS-KNOWN,RATIO}}(\boldsymbol{x}) \;=\; \frac{\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(\boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}}{\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}}. \tag{288}$$

These subspace detectors are desirable because they are UMP detectors. That is, they are provably the optimal detectors for detecting an unknown signal, regardless of its strength, in a known interference environment. Indeed, these detectors should be optimal for all values of A whenever $\mathbf{R}_\phi$ is known.

In order to make these optimal detectors usable, however, some choice must be made for θ and τ. Since the detectors above were created by maximizing the likelihood ratio, an appropriate set of estimators would be the maximum likelihood estimators. These estimators are the ones that, in this case, also maximize the detection statistics. Thus, a search must be made over all θ and τ to maximize the detection statistic, thus creating estimates of these parameters. Once τ and θ have been estimated, these detectors may be used. In reality, however, searching over θ and τ produces new detectors,

$$y_{\text{SUBS}}(\boldsymbol{x}) \;=\; \max_\phi \boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(\boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}, \text{ and} \tag{289}$$

$$y_{\text{SUBS-RATIO}}(\boldsymbol{x}) \;=\; \max_\phi \frac{\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \mathbf{R}_\phi \boldsymbol{\Psi} \left(\boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}}{\boldsymbol{x}^\dagger \mathbf{R}_n^{-1} \boldsymbol{x}}. \tag{290}$$
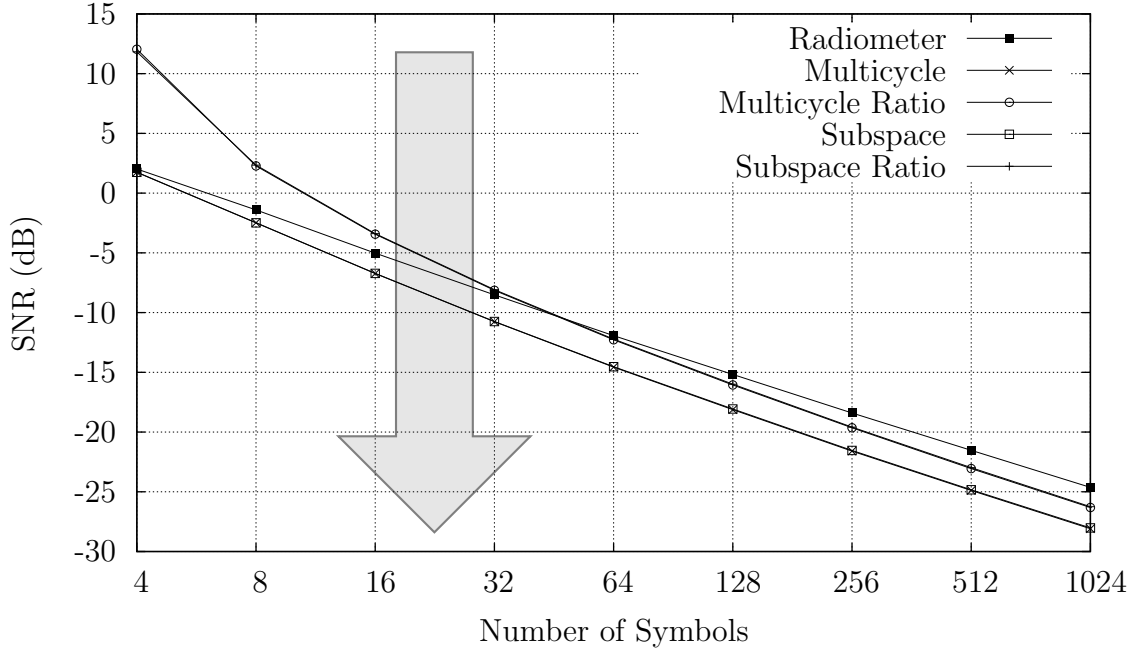
Figure 69. Subspace Detector Performance in White Noise

These new detectors are applicable to the problem of detecting an unknown signal in known noise, yet they are no longer the optimal detectors presented by Scharf and Friedlander.

Further, given that the detectors developed in Sec. 3.4 were developed under the assumption that the data vector, $\boldsymbol{d}$, was Gaussian, and given that subspace detectors have no such caveats, one might hypothesize that *uniformly* most powerful subspace detectors would outperform the *locally* most powerful cyclostationary detectors for small numbers of symbols where the Gaussian assumption is most suspect. Therefore, Fig. 69 plots the SNR required to detect a signal, with $P_{\mathrm{D}} = 0.9$ and $P_{\mathrm{FA}} = 0.1$, as a function of the number of symbols. Unlike previous detection performance diagrams, the best detector in this diagram is the one with the lowest required SNR.

From this diagram, it is difficult to tell which detector is better. Indeed, for large numbers of symbols, $N_s > 16$, the multicycle and multicycle ratio detectors

227

1.8
1.9
-0.15
-0.10
-0.05

0.12
0.14

0.16
0.18

0.22
0.24

0.26
0.28

0.45
0.50

Pre–Detection Filter

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Subspace
Low–SNR

Zero Cycle

Carrier

Symbol Rate

0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40

Normalized Frequency (Cycles / Sample)
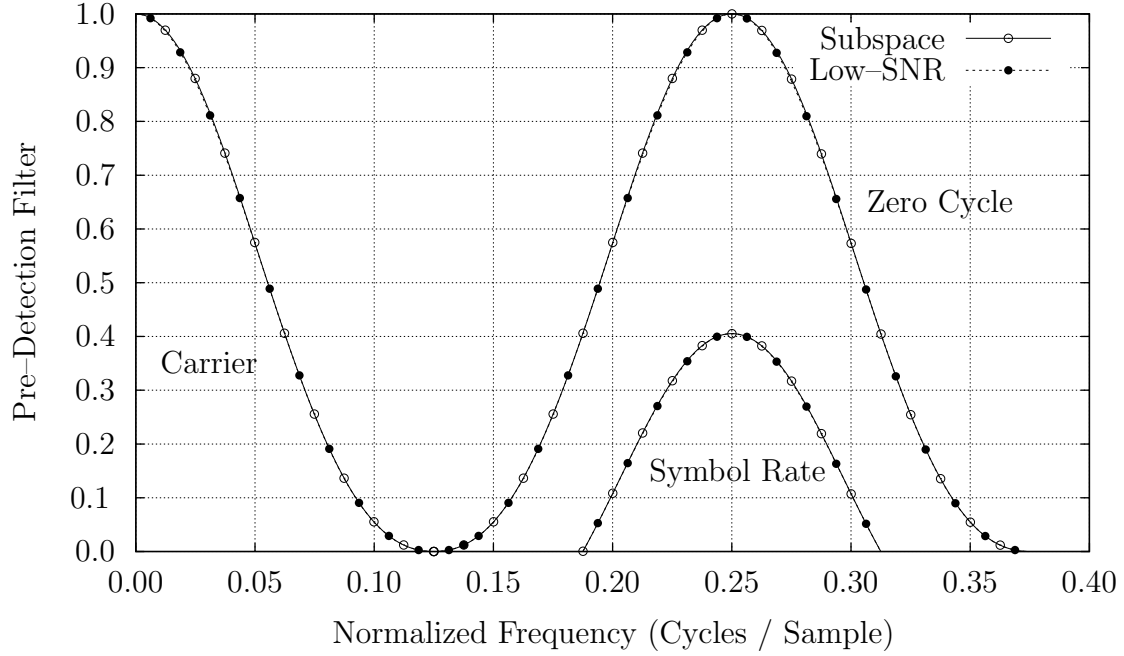
Figure 70.   Detection Functions for a White Noise Signal

differ only in the third decimal place from their subspace counterparts. In this third decimal place, however, the optimal detectors are the locally most powerful cyclostationary detectors.

Fig. 69, however, does not tell the whole story. Comparing the matrix internal to the subspace detector under white noise conditions,

$$\mathbf{Q}_{\text{SUBS}} \;=\; \mathbf{R}_\phi \boldsymbol{\Psi} \left(\boldsymbol{\Psi}^\dagger \mathbf{R}_n^{-1} \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger, \tag{291}$$

with the matrix internal to the multicycle detector,

$$\mathbf{Q}_{\text{MCYC}} \;=\; \mathbf{R}_\phi \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger, \tag{292}$$

reveals no significant differences. For reference, the diagonal and off diagonals of these matrices, called detection functions here, are plotted in Fig. 70, and separated

228
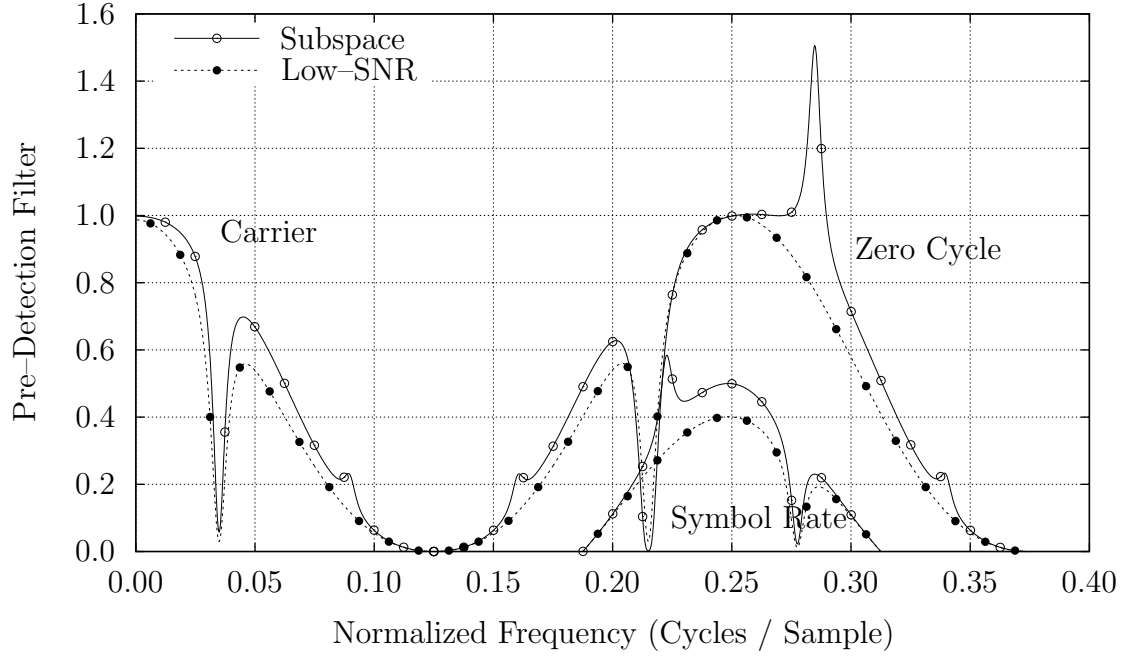
Figure 71.   Detection Functions for a Colored Noise Signal

according to cycle frequencies that they correspond to. Since these internal detection functions are near identical, this test may be ruled inconclusive.

If, on the other hand, the noise were colored instead of white, then these matrix diagonals would be quite different from each other. Using the colored noise spike introduced in Sec. 4.2.1, the colored noise detection functions are shown in Fig. 71. In this case the equalizer within the subspace detector, $\left(\mathbf{\Psi}^{\dagger}\mathbf{R}_n^{-1}\mathbf{\Psi}\right)$, significantly differentiates it from the multicycle detector.

Yet, in terms of performance, there is little measurable difference between their performance under these colored noise conditions, shown in Fig. 72, and their performance under the white noise conditions in Fig. 69. As before, the optimal detector is the multicycle detector. Also, as before, the difference between the detection SNRs of these two detectors is in the third decimal place once $N_s > 16$. Unlike before, however, the subspace detectors clearly outperform the multicycle detectors when $N_s \leq 10$.

−40
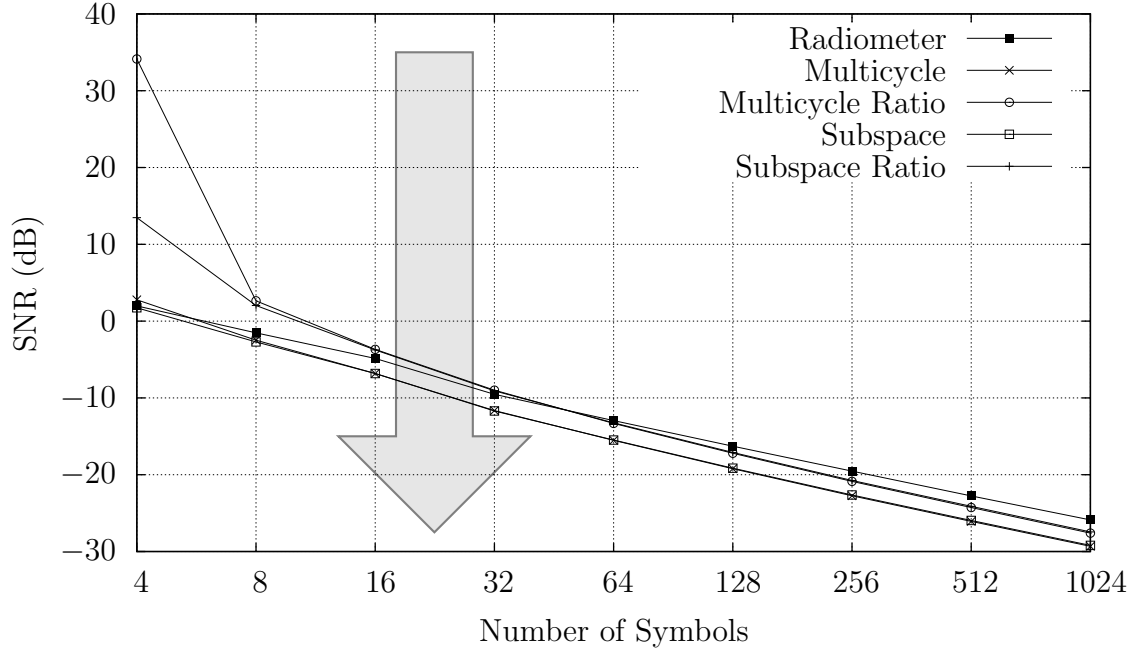
Figure 72.   Subspace Detector Performance in Colored Noise

These colored noise results validate the earlier hypothesis: The optimal detector for very small numbers of symbols is indeed the subspace detector. Under these circumstances, approximating $\boldsymbol{d}$ as a Gaussian is questionable. The fact that the multicycle detectors work so well for large numbers of symbols can easily be attributed to the fact that $\boldsymbol{d}$ can be validly modeled as a Gaussian random vector in frequency when the number of symbols is large. This Gaussian data model, while not perfect, is better than the subspace model in which no information at all is known about $\boldsymbol{d}$. Therefore, despite the fact that subspace detectors are provably the optimal detectors for detecting an unknown signal in a known subspace, they are not the optimal detectors for detecting PAM waveforms of moderate duration, $N_s > 10$.

Appendix J. Deriving the Cramér–Rao Bounds

This appendix presents much of the mathematical work behind the Cramér–Rao bound estimates presented within the text. As with deriving the Cyclic Ratio detector, this section is intended to be as complete as possible without skipping any steps. The unfortunately result is that the accompanying text may be somewhat terse.

Prior to beginning, two fundamental definitions are required. These are,

$$
\mathbf{F}_\delta \mathbf{R}_\phi \triangleq -\frac{1}{j2\pi} \frac{\partial}{\partial \tau} \mathbf{R}_\phi = \begin{bmatrix} (f_i - f_c) & & & 0 \\ & \ddots & & \\ & & \left(f_i + \frac{1}{T_s} - f_c\right) & \\ 0 & & & \ddots \end{bmatrix} \mathbf{R}_\phi
$$
(293)

$$
\text{and } \mathbf{F}_\Delta \mathbf{D}_k \triangleq -\frac{1}{j2\pi} \frac{\partial}{\partial \tau_{dk}} \mathbf{D}_k = \begin{bmatrix} f_i & 0 & & \\ & \ddots & & \\ & & f_i + \frac{1}{T_s} & \\ 0 & & & \ddots \end{bmatrix} \mathbf{D}_k = (\mathbf{F}_\delta + f_c \mathbf{I}) \, \mathbf{D}_k.
$$
(294)

Using these definitions, the derivatives of $\mathbf{R}_\phi$ and $\mathbf{D}_k$ can be taken quickly and easily,

$$
\begin{aligned}
\frac{\partial}{\partial \tau} \mathbf{R}_\phi &= -j2\pi \mathbf{F}_\delta \mathbf{R}_\phi, & \frac{\partial}{\partial \tau} \mathbf{R}_\phi^\dagger &= j2\pi \mathbf{F}_\delta \mathbf{R}_\phi^\dagger, \\
\frac{\partial}{\partial \tau} \mathbf{D}_k &= -j2\pi \mathbf{F}_\Delta \mathbf{D}_k, \text{ and} & \frac{\partial}{\partial \tau} \mathbf{D}_k^\dagger &= j2\pi \mathbf{F}_\Delta \mathbf{D}_k^\dagger
\end{aligned}
$$

Using these formulas, it is possible to keep the expressions for the Cramér–Rao bound simple, and in matrix form.

The Cramér–Rao bound for a vector of parameters, such as $\boldsymbol{\theta} \triangleq \begin{bmatrix} \tau & A & \boldsymbol{d} \end{bmatrix}^{T}$, is well known, and given by [53]

$$\mathcal{E}_{\boldsymbol{x}} \left\{ -\nabla_{\boldsymbol{\theta}}^{2} \mathcal{L} \right\}^{-1} . \tag{295}$$

The expression, $\mathcal{E}_{\boldsymbol{x}} \left\{ -\nabla_{\boldsymbol{\theta}}^{2} \mathcal{L} \right\}$ is known as the Fisher Information Matrix (FIM). The first step to evaluating any Cramér–Rao bound is to evaluate this matrix. Only then can it be inverted.

In this case, however, a related quantity will be evaluated leading to the *Modified Cramér–Rao Bound* (MCRB). Instead of taking the expectation over $\boldsymbol{x}$, the expectation will be taken over both $\boldsymbol{x}$ and $\boldsymbol{d}$, resulting in a modified FIM. The resulting bound is provably lower than the true CRB, and easier to evaluate [9].

The purpose of this appendix is to first evaluate the modified FIM, and second to invert it to get expressions for the Cramér–Rao bound. To do this, this appendix is broken into two sections. The first deals with the Cramér–Rao bounds for the unknown parameters on a single sensor. The second section deals with evaluating the FIM for a multiple sensor case. This latter section will conclude with Cramér–Rao bound for the two sensor TDOA problem when τ must be estimated as well.

J.1 Single Sensor Bounds

Starting with the second derivative of $\mathcal{L}$ for τ, we get the following expressions,

$$\mathcal{L} = -\frac{mN_f + N_f}{2}\ln(2\pi) - \frac{1}{2}\ln\det\mathbf{R}_n - \frac{1}{2}\ln\det\mathbf{R}_d \tag{296}$$

$$-\frac{1}{2}\left(\boldsymbol{x} - \frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right)^\dagger \mathbf{R}_n^{-1}\left(\boldsymbol{x} - \frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right) - \frac{1}{2}\boldsymbol{d}\mathbf{R}_d^{-1}\boldsymbol{d} \tag{297}$$

$$\frac{\partial\mathcal{L}}{\partial\tau} = \frac{A}{4}\frac{\partial}{\partial\tau}\left[\boldsymbol{x}^\dagger\mathbf{R}_n^{-1}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} + \boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\boldsymbol{x}\right] \tag{298}$$

$$= j2\pi\frac{A}{4}\left[-\boldsymbol{x}^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} + \boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{F}_\delta\mathbf{R}_n^{-1}\boldsymbol{x}\right] \tag{299}$$

$$\frac{\partial^2\mathcal{L}}{\partial\tau^2} = -4\pi^2\frac{A}{4}\left[\boldsymbol{x}^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta^2\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} + \boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{F}_\delta^2\mathbf{R}_n^{-1}\boldsymbol{x}\right] \tag{300}$$

$$= -2\pi^2 A\Re\left\{\boldsymbol{x}^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta^2\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right\} \tag{301}$$

Placing these expressions into the expression for the FIM provides what could be the MCRB for τ (once inverted),

$$\mathcal{E}\left\{-\frac{\partial^2\mathcal{L}}{\partial\tau^2}\right\} = A^2\pi^2\mathcal{E}\left\{\boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta^2\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right\} \tag{302}$$

$$= A^2\pi^2\mathcal{E}\left\{\boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta^2\boldsymbol{\Psi}\boldsymbol{d}\right\} \tag{303}$$

$$= A^2\pi^2\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta^2\right\}. \tag{304}$$

The problem with stopping here is that τ is only one of the unknowns in a single sensor problem. $\boldsymbol{d}$ and A must be estimated as well.

Therefore, we turn our attention next to the unknown data, $\boldsymbol{d}$,

$$\nabla_{\boldsymbol{d}^\dagger}\mathcal{L} = \frac{A}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\boldsymbol{x} - \left(\frac{A^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)\boldsymbol{d} \tag{305}$$

$$\mathcal{E}\left\{-\nabla_{\boldsymbol{d}}^2\mathcal{L}\right\} = \frac{A^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1} \tag{306}$$

$$\frac{\partial}{\partial\tau}\nabla_{\boldsymbol{d}^\dagger}\mathcal{L} = j\left(2\pi\right)\frac{A}{2}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{F}_\delta\mathbf{R}_n^{-1}\boldsymbol{x} \tag{307}$$

$$\mathcal{E}\left\{-\frac{\partial\nabla_{\boldsymbol{d}^\dagger}\mathcal{L}}{\partial\tau}\right\} = -j\left(2\pi\right)\frac{A}{2}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{F}_\delta\mathbf{R}_n^{-1}\mathcal{E}\left\{\boldsymbol{x}\right\} \tag{308}$$

$$= -j\left(2\pi\right)\frac{A}{2}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{F}_\delta\mathbf{R}_n^{-1}\frac{A}{2}\mathbf{R}_\phi\boldsymbol{\Psi}\mathcal{E}\left\{\boldsymbol{d}\right\} \tag{309}$$

$$= \mathbf{0}. \tag{310}$$

Finally, we turn our attention to the unknown amplitude,

$$\frac{\partial^2\mathcal{L}}{\partial A\partial\tau} = j\pi\frac{1}{2}\left[-\boldsymbol{x}^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} + \boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{F}_\delta\mathbf{R}_n^{-1}\boldsymbol{x}\right] \tag{311}$$

$$\mathcal{E}\left\{-\frac{\partial^2\mathcal{L}}{\partial A\partial\tau}\right\} = j\pi\frac{A}{4}\mathcal{E}_{\boldsymbol{d}}\left\{-\boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} + \boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right\} \tag{312}$$

$$= 0 \tag{313}$$

$$\frac{\partial\mathcal{L}}{\partial A} = \frac{1}{4}\left[\boldsymbol{x}^\dagger\mathbf{R}_n^{-1}\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} + \boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\boldsymbol{x}\right] - \frac{A}{4}\boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\boldsymbol{\Psi}\boldsymbol{d} \tag{314}$$

$$\mathcal{E}\left\{-\frac{\partial^2\mathcal{L}}{\partial A^2}\right\} = \frac{1}{4}\text{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\right\} \tag{315}$$

$$\frac{\partial\nabla_{\boldsymbol{d}^\dagger}\mathcal{L}}{\partial A} = \frac{1}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{R}_n^{-1}\boldsymbol{x} - \frac{A}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\boldsymbol{\Psi}\boldsymbol{d} \tag{316}$$

$$\mathcal{E}\left\{-\frac{\partial\nabla_{\boldsymbol{d}^\dagger}\mathcal{L}}{\partial A}\right\} = \mathbf{0} \tag{317}$$

234

So there are no interdependencies and the Cramér–Rao bounds are quickly calculated,

$$
MCRB\left(\tau\right) = \frac{1}{A^2\pi^2} \frac{1}{\operatorname{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\mathbf{F}_\delta^2\right\}} \tag{318}
$$

$$
MCRB\left(A\right) = \frac{4}{\operatorname{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\right\}} \tag{319}
$$

$$
\text{and } MCRB\left(\boldsymbol{d}\right) = \left(\frac{A^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_n^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)^{-1} \tag{320}
$$

Two things commend these bounds to us. The first is that the MCRB for τ is already known for the baseband, $f_c = 0$, white noise case, $\mathbf{R}_n = N_o\mathbf{I}$, and this expression for the MCRB is consistent with that known result. [9] The second is that the bound for $\boldsymbol{d}$ is discussed by Berger and Tufts in [2] under the case of known time delay. Likewise, this expression is also consistent with Berger and Tufts work. These two known results commend this method, prior to extending it to multiple sensor evaluations.

J.2   Multiple Sensor Bounds

Using the assumption that the interference is uncorrelated between each sensor, the likelihood function is easy to write,

$$
\mathcal{L} = -\frac{mN_fM + N_f}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\det\mathbf{R}_d - \frac{1}{2}\sum_{k=0}^{M-1}\ln\det\mathbf{R}_{nk} \tag{321}
$$

$$
-\frac{1}{2}\boldsymbol{d}\mathbf{R}_d^{-1}\boldsymbol{d} - \frac{1}{2}\sum_{k=0}^{M-1}\left(\boldsymbol{x}_k - e^{j\theta_{dk}}\frac{A_k}{2}\mathbf{D}_k\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right)^\dagger \mathbf{R}_{nk}^{-1}\left(\boldsymbol{x}_k - e^{j\theta_{dk}}\frac{A_k}{2}\mathbf{D}_k\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right)
$$

Given this likelihood function, derivatives proceed as before. First, for the unknown data vector,

$$\nabla_{\boldsymbol{d}^\dagger}\mathcal{L} = \sum_{k=0}^{M-1} e^{-j\theta_{dk}}\frac{A_k}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{D}_k^\dagger\mathbf{R}_{nk}^{-1}\boldsymbol{x}_k - \left(\sum_{k=0}^{M-1}\frac{A_k^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_{nk}^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}\right)\boldsymbol{d} \tag{322}$$

$$\mathcal{E}\left\{-\nabla_{\boldsymbol{d}}^2\mathcal{L}\right\} = \sum_{k=0}^{M-1}\frac{A_k^2}{4}\boldsymbol{\Psi}^\dagger\mathbf{R}_{nk}^{-1}\boldsymbol{\Psi} + \mathbf{R}_d^{-1}. \tag{323}$$

The next parameter is the unknown amplitude. In this case, there is one unknown amplitude on each sensor, and the effect that unknown amplitude has on the data vector needs to be evaluated as well.

$$\mathcal{E}\left\{-\frac{\partial^2\mathcal{L}}{\partial A_i^2}\right\} = \frac{1}{4}\mathrm{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{ni}^{-1}\right\} \tag{324}$$

$$\mathcal{E}\left\{-\frac{\partial\nabla_{\boldsymbol{d}^\dagger}\mathcal{L}}{\partial A_k}\right\} = \mathbf{0} \tag{325}$$

$$\mathcal{E}\left\{-\frac{\partial^2\mathcal{L}}{\partial A_i\partial A_k}\right\}_{k\neq i} = 0 \tag{326}$$

Then we turn to τ and τ_{di}, as each of these need to be estimated,

$$\frac{\partial \mathcal{L}}{\partial \tau} = \frac{1}{2} \sum_{k=0}^{M-1} \frac{A_k}{2} \left[\begin{array}{l} - \quad j2\pi e^{j\theta_{dk}} \boldsymbol{x}_k^\dagger \mathbf{R}_{nk}^{-1} \mathbf{F}_\delta \mathbf{D}_k \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \\ + \quad j2\pi e^{-j\theta_{dk}} \boldsymbol{d}^\dagger \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{D}_k^\dagger \mathbf{F}_\delta \mathbf{R}_{nk}^{-1} \boldsymbol{x}_k \end{array} \right] \tag{327}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \tau^2} = \sum_{k=0}^{M-1} \frac{A_k}{2} \Re \left\{ -4\pi^2 e^{j\theta_{dk}} \boldsymbol{x}_k^\dagger \mathbf{R}_{nk}^{-1} \mathbf{F}_\delta^2 \mathbf{D}_k \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right\} \tag{328}$$

$$\mathcal{E}\left\{ -\frac{\partial^2 \mathcal{L}}{\partial \tau^2} \right\} = \sum_{k=0}^{M-1} \frac{A_k}{2} \mathcal{E}_{\boldsymbol{d}} \left\{ \Re \left\{ \begin{array}{c} 4\pi^2 e^{j\theta_{dk}} \left(e^{j\theta_{dk}} \frac{A_k}{2} \mathbf{D}_k \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \right)^\dagger \\ \times \mathbf{R}_{nk}^{-1} \mathbf{F}_\delta^2 \mathbf{D}_k \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \end{array} \right\} \right\} \tag{329}$$

$$= \sum_{k=0}^{M-1} A_k^2 \pi^2 \mathrm{tr} \left\{ \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_{nk}^{-1} \mathbf{F}_\delta^2 \right\} \tag{330}$$

$$\frac{\partial \mathcal{L}}{\partial \tau_{di}} = \frac{A_i}{4} \left[\begin{array}{l} - \quad j2\pi e^{j\theta_{di}} \boldsymbol{x}_i^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\Delta \mathbf{D}_i \mathbf{R}_\phi \boldsymbol{\Psi} \boldsymbol{d} \\ + \quad j2\pi e^{-j\theta_{di}} \boldsymbol{d}^\dagger \boldsymbol{\Psi}^\dagger \mathbf{R}_\phi^\dagger \mathbf{D}_i^\dagger \mathbf{F}_\Delta \mathbf{R}_{ni}^{-1} \boldsymbol{x}_i \end{array} \right] \tag{331}$$

$$\mathcal{E}\left\{ -\frac{\partial^2 \mathcal{L}}{\partial \tau_{di}^2} \right\} = A_i^2 \pi^2 \mathrm{tr} \left\{ \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\Delta^2 \right\} \tag{332}$$

$$\mathcal{E}\left\{ -\frac{\partial^2 \mathcal{L}}{\partial \tau_{di} \partial \tau} \right\} = A_i^2 \pi^2 \mathrm{tr} \left\{ \boldsymbol{\Psi} \mathbf{R}_d \boldsymbol{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\delta \mathbf{F}_\Delta \right\} \tag{333}$$

This leaves one final parameter of interest that may need to be estimated, θ_{di}.

$$\frac{\partial \mathcal{L}}{\partial \theta_{di}} = je^{j\theta_{di}}\frac{A_i}{4}\boldsymbol{x}_i^\dagger \mathbf{R}_{ni}^{-1}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} - je^{-j\theta_{di}}\frac{A_i}{4}\boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{D}_i^\dagger\mathbf{R}_{ni}^{-1}\boldsymbol{x}_i \quad (334)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \theta_{di}^2} = -e^{j\theta_{di}}\frac{A_i}{4}\boldsymbol{x}_i^\dagger \mathbf{R}_{ni}^{-1}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} - e^{-j\theta_{di}}\frac{A_i}{4}\boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{D}_i^\dagger\mathbf{R}_{ni}^{-1}\boldsymbol{x}_i \quad (335)$$

$$= -\frac{A_i}{2}\Re\left\{e^{j\theta_{di}}\boldsymbol{x}_i^\dagger \mathbf{R}_{ni}^{-1}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right\} \quad (336)$$

$$\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{di}^2}\right\} = \frac{A_i}{2}\Re\left\{e^{j\theta_{di}}\left(\frac{A_i}{2}e^{j\theta_i}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right)^\dagger \mathbf{R}_{ni}^{-1}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right\} \quad (337)$$

$$= \frac{A_i^2}{4}\mathrm{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{ni}^{-1}\right\} \quad (338)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \theta_{di}\partial \tau} = \frac{A_i}{2}\left[\pi e^{j\theta_{di}}\boldsymbol{x}_i^\dagger \mathbf{R}_{ni}^{-1}\mathbf{F}_\delta\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d} + \pi e^{-j\theta_{di}}\boldsymbol{d}^\dagger\boldsymbol{\Psi}^\dagger\mathbf{R}_\phi^\dagger\mathbf{F}_\delta\mathbf{D}_i^\dagger\mathbf{R}_{ni}^{-1}\boldsymbol{x}_i\right]$$

$$\quad (339)$$

$$= A_i\Re\left\{\pi e^{j\theta_{di}}\boldsymbol{x}_i^\dagger \mathbf{R}_{ni}^{-1}\mathbf{F}_\delta\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right\} \quad (340)$$

$$\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{di}\partial \tau}\right\} = -A_i\mathcal{E}_{\boldsymbol{d}}\left\{\Re\left\{\pi e^{j\theta_{di}}\left(e^{j\theta_{di}}\frac{A_i}{2}\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right)^\dagger \mathbf{R}_{ni}^{-1}\mathbf{F}_\delta\mathbf{D}_i\mathbf{R}_\phi\boldsymbol{\Psi}\boldsymbol{d}\right\}\right\}$$

$$\quad (341)$$

$$= -\frac{A_i^2\pi}{2}\mathrm{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{ni}^{-1}\mathbf{F}_\delta\right\} \quad (342)$$

$$\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{di}\partial \tau_{di}}\right\} = -\frac{A_i^2\pi}{2}\mathrm{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{ni}^{-1}\mathbf{F}_\Delta\right\} \quad (343)$$

From here, two MCRBs are immediately apparent. These are the MCRB for A_i and the MCRB for $\boldsymbol{d}$, since these two bounds are independent. That is, the inverse is easily applied,

$$MCRB\left(A_i\right) = \frac{4}{\mathrm{tr}\left\{\boldsymbol{\Psi}\mathbf{R}_d\boldsymbol{\Psi}^\dagger\mathbf{R}_{ni}^{-1}\right\}} \quad (344)$$

$$MCRB\left(\boldsymbol{d}\right) = \left(\mathbf{R}_d^{-1} + \frac{1}{4}\sum_{k=0}^{M-1}A_k^2\boldsymbol{\Psi}\mathbf{R}_{nk}^{-1}\boldsymbol{\Psi}^\dagger\right)^{-1} \quad (345)$$

The difficult part is to find an analytical expression for the MCRBs of τ, θ_{di}, and τ_{di}. Since this involves inverting an $M \times M$ matrix, analytical results are difficult to obtain. Instead, the previous results are summarized and then a two sensor problem is considered.

Recalling from the above equations, and evaluating when $\mathbf{R}_d = N_s \mathbf{I}$, the following modified FIM entries are known:

$$\mathcal{E}\left\{ -\frac{\partial^2 \mathcal{L}}{\partial \tau^2} \right\} = \sum_{k=0}^{M-1} A_k^2 \pi^2 \mathrm{tr}\left\{ \mathbf{\Psi} \mathbf{R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{nk}^{-1} \mathbf{F}_\delta^2 \right\} \tag{346}$$

$$= \pi^2 \sum_{k=0}^{M-1} A_k^2 \sum_f \frac{|\Psi(f - f_c)|^2}{T_s S_{nk}(f)} (f - f_c)^2 \tag{347}$$

$$\mathcal{E}\left\{ -\frac{\partial^2 \mathcal{L}}{\partial \tau_{di}^2} \right\} = A_i^2 \pi^2 \mathrm{tr}\left\{ \mathbf{\Psi} \mathbf{R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\Delta^2 \right\} \tag{348}$$

$$= \pi^2 A_i^2 \sum_f \frac{|\Psi(f - f_c)|^2}{T_s S_{ni}(f)} f^2 \tag{349}$$

$$\mathcal{E}\left\{ -\frac{\partial^2 \mathcal{L}}{\partial \tau_{di} \partial \tau} \right\} = A_i^2 \pi^2 \mathrm{tr}\left\{ \mathbf{\Psi} \mathbf{R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\delta \mathbf{F}_\Delta \right\} \tag{350}$$

$$= \pi^2 A_k^2 \sum_f \frac{|\Psi(f - f_c)|^2}{T_s S_{ni}(f)} f(f - f_c) \tag{351}$$

$$\mathcal{E}\left\{ -\frac{\partial^2 \mathcal{L}}{\partial \theta_{di}^2} \right\} = \frac{A_i^2}{4} \mathrm{tr}\left\{ \mathbf{\Psi} \mathbf{R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \right\} \tag{352}$$

$$= \frac{A_i^2}{4} \sum_f \frac{|\Psi(f - f_c)|^2}{T_s S_{ni}(f)} \tag{353}$$

$$\mathcal{E}\left\{ -\frac{\partial^2 \mathcal{L}}{\partial \theta_{di} \partial \tau} \right\} = -\frac{A_i^2 \pi}{2} \mathrm{tr}\left\{ \mathbf{\Psi} \mathbf{R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\delta \right\} \tag{354}$$

$$= -\frac{A_i^2 \pi}{2} \sum_f \frac{|\Psi(f - f_c)|^2}{T_s S_{ni}(f)} (f - f_c) \tag{355}$$

$$\mathcal{E}\left\{ -\frac{\partial^2 \mathcal{L}}{\partial \theta_{di} \partial \tau_{di}} \right\} = -\frac{A_i^2 \pi}{2} \mathrm{tr}\left\{ \mathbf{\Psi} \mathbf{R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{ni}^{-1} \mathbf{F}_\Delta \right\} \tag{356}$$

$$= -\frac{A_i^2 \pi}{2} \sum_f \frac{|\Psi(f - f_c)|^2}{T_s S_{ni}(f)} f \tag{357}$$

$$
FIM = \begin{pmatrix}
\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau^2}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{d1} \partial \tau}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d1} \partial \tau}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{d2} \partial \tau}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d2} \partial \tau}\right\} & \cdots \\
\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{d1} \partial \tau}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{d1}^2}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d1} \partial \tau_{d1}}\right\} & 0 & 0 & \cdots \\
\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d1} \partial \tau}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d1} \partial \tau_{d1}}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d1}^2}\right\} & 0 & 0 & \cdots \\
\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{d2} \partial \tau}\right\} & 0 & 0 & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{d2}^2}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d2} \partial \tau_{d2}}\right\} & \cdots \\
\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d2} \partial \tau}\right\} & 0 & 0 & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d2} \partial \tau_{d2}}\right\} & \mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \theta_{d2}^2}\right\} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\tag{358}
$$

In the particular case of a two sensor problem where $\theta_{d1} = 0$, the MCRB for τ_{d1} is given by,

$$
\begin{aligned}
MCRB_{bp}(\tau_{d1}) &= \frac{1}{\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{d1}^2}\right\} - \frac{\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau_{d1} \partial \tau}\right\}^2}{\mathcal{E}\left\{-\frac{\partial^2 \mathcal{L}}{\partial \tau^2}\right\}}} \\
&= \frac{1}{\pi^2 A_1^2 \mathrm{tr}\left\{\mathbf{\Psi R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{n1}^{-1} \mathbf{F}_\delta^2\right\} - \pi^2 A_1^2 \frac{A_1^2 \mathrm{tr}\left\{\mathbf{\Psi R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{n1}^{-1} \mathbf{F}_\delta \mathbf{F}_\Delta\right\}^2}{A_0^2 \mathrm{tr}\left\{\mathbf{\Psi R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{n0}^{-1} \mathbf{F}_\delta^2\right\} + A_1^2 \mathrm{tr}\left\{\mathbf{\Psi R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{n1}^{-1} \mathbf{F}_\delta^2\right\}}}
\end{aligned}
\tag{359}
$$

This is the function evaluated for the colored noise TDOA estimation tests given in Figs. 43 and 45.

A similar function, appropriate for baseband TDOA estimation, is shown in Figs. 37 and 38. This function is given by,

$$
MCRB_{bb}(\tau_{d1}) = \frac{1}{4\pi^2 A_1^2 \mathrm{tr}\left\{\mathbf{\Psi R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{n1}^{-1} \mathbf{F}_\delta^2\right\} - 4\pi^2 \frac{A_1^4 \mathrm{tr}\left\{\mathbf{\Psi R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{n1}^{-1} \mathbf{F}_\delta^2\right\}^2}{A_0^2 \mathrm{tr}\left\{\mathbf{\Psi R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{n0}^{-1} \mathbf{F}_\delta^2\right\} + A_1^2 \mathrm{tr}\left\{\mathbf{\Psi R}_d \mathbf{\Psi}^\dagger \mathbf{R}_{n1}^{-1} \mathbf{F}_\delta^2\right\}}}.
\tag{360}
$$

The extra factor of four in this equation is a result of the fact that a baseband signal has a scale factor of A_i, and not $\frac{A_i}{2}$.

Glossary

Autocorrelation function The autocorrelation function of a random process, $x(t)$, is defined by,

$$R_x(t, \Delta t) \triangleq \mathcal{E}\left\{ x^*\left(t - \frac{\Delta t}{2}\right) x\left(t + \frac{\Delta t}{2}\right) \right\}. \tag{361}$$

If the random process is stationary, this autocorrelation function is a function of time difference, Δt, only. If, instead, the function is cyclostationary, its autocorrelation function admits a Fourier series expansion,

$$R_x(t, \Delta t) = \sum_\alpha R_x^\alpha(\Delta t) \, e^{j2\pi\alpha t}, \tag{362}$$

where α ranges over all of the cycle frequencies found in the signal.

Bandlimited A signal is bandlimited if its Fourier transform is zero or near zero outside of a particular band of frequencies.

Bandpass A signal having a non–zero carrier frequency, f_c, that is typically much greater than its bandwidth.

Bandwidth The width, in frequency, of the spectrum used by a particular signal.

Baseband A "signal whose spectrum extends from (or near) dc [sic] up to some finite value, usually less than a few megahertz" [58, p. 56].

Binary Phase Shift Keying A Binary Phase Shift Keyed (BPSK) system is one where the weights modifying the pulse function are plus or minus one, $d_n \in \{\pm 1\}$.

Burst A burst is a signal of finite duration having two properties. The first property is that a burst can be observed in its entirety. The second property is that it is

uncommon enough that the noise covariance may be measured when the burst is not present.

Colored Noise A stationary random process possessing some amount of correlation in time. This is equivalent to stating that its power spectral density is not constant.

Complex Signal A signal created by taking a complex–valued function, with real and imaginary components, and multiplying it by a complex exponential carrier. The real portion of the result is then transmitted.

Cramér–Rao Bound A lower bound on the error covariance matrix for any unbiased estimator of an unknown parameter [53].

Cross Spectral Density The cross spectral density between two signals, $x(t)$ and $y(t)$, is defined by,

$$S_{xy}(f) \triangleq \lim_{T \to \infty} \frac{1}{T} \mathcal{E}\left\{ X_T(f)^* Y_T(f) \right\}, \text{where} \tag{363}$$

$$X_T(f) \triangleq \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-j2\pi ft} dt. \tag{364}$$

Cyclostationary A random process is called *cyclostationary* if all the probability distribution functions describing the process are periodic or polyperiodic functions of time.

Discrete PSD The power spectral density of a discrete sequence, d_n, is defined by

$$S_d\left(e^{j2\pi f}\right) \triangleq \lim_{N_s \to \infty} \mathcal{E}\left\{ \frac{1}{N_s} \left| \sum_{n=0}^{N_s-1} d_n e^{-j2\pi fn} \right|^2 \right\}.$$

242

Expected Value The *expected value* or *mean* of a continuous random variable, $g(\boldsymbol{X})$, denoted by $\mathcal{E}\{g(\boldsymbol{X})\}$, is

$$\mathcal{E}\{g(\boldsymbol{X})\} = \int g(\boldsymbol{X}) f(\boldsymbol{x}) \, d\boldsymbol{x}, \tag{365}$$

where $f(\boldsymbol{x})$ is the probability density function for the random variable $\boldsymbol{x}$ [7, p. 55]. This definition assumes, of course, that the integral exists and is finite.

Field A *field* $\mathbb{F}$ is a set of numbers together with certain properties. In particular, if x, y, and z are arbitray elements in $\mathbb{F}$, then [31, p 1–2]

1. addition is commutative, $x + y = y + x$;
2. addition is associative, $x + (y + z) = (x + y) + z$;
3. there is a unique element 0 (zero) in $\mathbb{F}$ such that $x + 0 = x$;
4. a unique element, $(-x)$ corresponds to x such that $x + (-x) = 0$;
5. multiplication is commutative, $xy = yx$;
6. and multiplication is associative, $x(yz) = (xy)z$.

The two fields of particular interest to this research are the set of real numbers, $\mathbb{R}$, and the set of complex numbers, $\mathbb{C}$.

Fisher Information Matrix If $f_{\boldsymbol{\phi}}(\boldsymbol{x})$ is a probability density function parameterized by the vector, $\boldsymbol{\phi}$, then the Fisher Information Matrix (FIM) is given by the expected value of the negative of the second gradient of the log of $f_{\boldsymbol{\phi}}(\boldsymbol{x})$,

$$\mathbf{J} \triangleq \mathcal{E}\left\{-\nabla_{\boldsymbol{\phi}}^2 \ln f_{\boldsymbol{\phi}}(\boldsymbol{x})\right\}. \tag{366}$$

In particular, the elements within this matrix are given by,

$$\mathbf{J}_{ij} = \mathcal{E}\left\{-\frac{\partial^2}{\partial \phi_i \partial \phi_j} \ln f_{\boldsymbol{\phi}}(\boldsymbol{x})\right\}. \tag{367}$$

The diagonal elements of $\mathbf{J}^{-1}$ correspond to the Cramér–Rao bounds on the parameters within $\boldsymbol{\phi}$.

Gaussian Process A random process, $X(t)$, is considered *Gaussian*, if for every set of unique time instants, t_1, t_2, ..., t_N, the vector

$$\boldsymbol{x} = \left[\begin{array}{cccc} X(t_1) & X(t_2) & \cdots & X(t_N) \end{array} \right]^T \tag{368}$$

is a multivariate Gaussian.

Gaussian Random Variable A random variable, $X \in \mathbb{R}$ is said to be Gaussian, or to have a Gaussian distribution, if the probability that X lies within a range $(a, b) \subset \mathbb{R}$ is given by

$$P\left[a < X < b\right] = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx, \tag{369}$$

where μ and σ^2 are the mean and variance of X respectively. In this case, we say that X is Gaussian distributed with a mean μ and variance σ^2 and note this as, $X \sim \mathcal{N}(\mu, \sigma^2)$.

Geolocation The process of estimating the position of a transmitter (or receiver) on the earth's surface.

Intersymbol Interference Interference within the receiver caused by the tail of one pulse interfering with the detection of another pulse.

Linear Space From Hoffman and Kunze,

> A *vector space* (or *linear space*) consists of the following:
> 1. a field $\mathbb{F}$ of scalars;
> 2. a set $\mathcal{V}$ of objects, called vectors;
> 3. a rule (or operation), called vector addition, which associates with each pair of vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ in $\mathcal{V}$ a vector $\boldsymbol{\alpha} + \boldsymbol{\beta}$ in $\mathcal{V}$, called the sum of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, in such a way that
>   (a) addition is commutative, $\boldsymbol{\alpha} + \boldsymbol{\beta} = \boldsymbol{\beta} + \boldsymbol{\alpha}$;
>   (b) addition is associative, $\boldsymbol{\alpha} + (\boldsymbol{\beta} + \boldsymbol{\gamma}) = (\boldsymbol{\alpha} + \boldsymbol{\beta}) + \boldsymbol{\gamma}$;
>   (c) there is a unique vector $\boldsymbol{0}$ in $\mathcal{V}$, called the zero vector, such that $\boldsymbol{\alpha} + \boldsymbol{0} = \boldsymbol{\alpha}$ for all $\boldsymbol{\alpha}$ in $\mathcal{V}$;

      (d) for each vector $\boldsymbol{\alpha}$ in $\mathcal{V}$ there is a unique vector $-\boldsymbol{\alpha}$ in $\mathcal{V}$ such that $\boldsymbol{\alpha} + (-\boldsymbol{\alpha}) = 0$;

   4. a rule (or operation), called scalar multiplication, which associates with each scalar c in $\mathbb{F}$ and vector $\boldsymbol{\alpha}$ in $\mathcal{V}$ a vector $c\boldsymbol{\alpha}$ in $\mathcal{V}$, called the product of c and $\boldsymbol{\alpha}$, in such a way that

      (a) $1\boldsymbol{\alpha} = \boldsymbol{\alpha}$ for every $\boldsymbol{\alpha}$ in $\mathcal{V}$;

      (b) $(c_1 c_2)\,\boldsymbol{\alpha} = c_1\,(c_2\boldsymbol{\alpha})$;

      (c) $c\,(\boldsymbol{\alpha} + \boldsymbol{\beta}) = c\boldsymbol{\alpha} + c\boldsymbol{\beta}$;

      (d) $(c_1 + c_2)\,\boldsymbol{\alpha} = c_1\boldsymbol{\alpha} + c_2\boldsymbol{\alpha}$.

  [31, p. 28–20]

One vector space of particular interest to this research is the set of all real functions, $x\,(t)$, defined from $-\infty$ to ∞.

Linear Subspace Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$. A *linear subspace* of $\mathcal{V}$ is a subset $\mathcal{W}$ of $\mathcal{V}$ which is itself a vector space [31, p. 34].

The particular subspace of interest to this research is the subspace described by linear combinations of pulse functions, such as those given by Eqns. (39) and (40), since every possible received message may be represented as a linear combination of these functions.

Log Likelihood Function Given a probability density function for a random variable, $\boldsymbol{x}$, that is parameterized by θ, $f\,(\boldsymbol{x}|\theta)$, the likelihood function is the function of θ, $L\,(\theta)$, that results when $\boldsymbol{x}$ is fixed. The *log* likelihood function, $\mathcal{L}$ is simply the log of the likelihood function, $\mathcal{L} \triangleq \ln L\,(\theta)$.

Matched Filter A "linear filter designed to provide the maximum signal–to–noise power ratio at its output for a given transmitted symbol waveform" [58, p. 122].

Maximum Likelihood Estimate The maximum likelihood estimate of the parameter, θ, of a probability distribution function, $f\,(\boldsymbol{x}|\theta)$, is defined as the parameter that maximizes the likelihood function, $\mathcal{L}\,(\theta)$.

Multivariate Gaussian An N–dimensional vector $\boldsymbol{x}$ is said to be a multivariate Gaussian if the probability that it lies within a region $\mathcal{W}$ of the N–dimensional

vector space $\mathcal{V}$ is given by

$$P\left[\boldsymbol{x} \in \mathcal{W}\right] \;=\; \int_{\mathcal{W}} \left(2\pi\right)^{-\frac{N}{2}} \left|\mathbf{R}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\dagger}\mathbf{R}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right\} d\boldsymbol{x}. \quad (370)$$

In this case we say that $\boldsymbol{x}$ has a multivariate Gaussian distribution with mean, $\boldsymbol{\mu}$ and covariance matrix, $\mathbf{R}$, and note this as $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{\mu}, \mathbf{R}\right)$.

Normalized Frequency Frequency units for sampled data, counting the number of cycles per sample. The maximum frequency is $\frac{1}{2}$, or one cycle every two samples.

Nuisance Parameter A parameter that is present in a model, but "not of direct inferential interest." [7, p. 378] That is, the parameter needs to be known or estimated in order to work with the model, but knowing the parameter does not necessarily provide any practical value to the task at hand.

Nyquist Bandwidth The theoretical minimum system bandwidth needed to demodulate a signal at a particular rate.

Overwhitener A filter defined by applying a whitener twice, $H_{\text{ovw}}\left(f\right) = \frac{1}{S_n(f)}$.

Polyperiodic Consisting of multiple, possibly incommensurate, periods.

Power Spectral Density The power distribution of a signal in the frequency domain. For a signal $x\left(t\right)$, this function is defined as,

$$S_x\left(f\right) \;\triangleq\; \lim_{T\to\infty} \mathcal{E}\left\{\frac{1}{T}\left|\int_{-\frac{T}{2}}^{\frac{T}{2}} x\left(t\right) e^{-j2\pi ft} dt\right|^2\right\}.$$

From here, the total power in the signal is given by integrating this expression across all frequency,

$$P_{\text{TOTAL}} \;=\; \int_{-\infty}^{\infty} S_x\left(f\right) df. \quad (371)$$

246

Similar expressions are used to calculate the total power in a signal that falls within a particular frequency band.

Quadrature Phase Shift Keying A Quadrature Phase Shift Keyed (QPSK) system is one where the weights modifying the pulse function form the four corners of a square on the unit circle. For example, $d_n \in \{\pm 1, \pm j\}$.

Random Process A real valued random process, $X(t)$, is a measurable function that maps a real number, t, to a random value selected from a subset of real, or possibly complex, numbers.

Real Signal Used to describe a signal created by taking a real–valued function and multiplying it by a sine–wave carrier.

Signal to Noise Ratio The ratio of the energy in the signal to the energy in the noise (sum of squares). Since this is only defined over a particular band, it is defined here as the energy in the signal over the band $f_c - \frac{1}{T_s}$ to $f_c + \frac{1}{T_s}$.

Spectral Correlation Function A statistical measure of the correlation between two frequencies of a signal. If $x(t)$ is a random process, this function is defined as,

$$S_x^\alpha(f) \triangleq \lim_{T \to \infty} \mathcal{E}\left\{ \frac{1}{T} X_T\left(f - \frac{\alpha}{2}\right)^* X_T\left(f + \frac{\alpha}{2}\right) \right\}$$

where

$$X_T(f) \triangleq \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)\, e^{-j2\pi ft} dt.$$

This function may also be referred to as the *cyclic spectral density*.

Stationary "A random process $X(t)$ is called time stationary or *stationary in the strict sense* if all of the [probability] distribution functions describing the process are invariant under a translation of time" [57, p. 135].

Tapped Delay Line Any linear filter having an impulse response represented by a sum of scaled delta functions spaced at periodic intervals.

Trace The *trace* of a matrix $\mathbf{A}$, denoted $\mathrm{tr}\{\mathbf{A}\}$ "is defined to be the sum of the diagonal entries of $\mathbf{A}$" [33, p. 216].

White Noise A stationary random process having the property that every sample in time is statistically independent of every other is considered a white random process. A consequence of this definition is that the power spectral density of this process is a constant. Since this is physically impossible, the phrase is more often used to refer to noise that is white across some bandwidth of interest.

Whitener A filter designed to produce white noise from a given colored noise input. (See Eqn. (14) on page 21.)

Wide–sense Stationarity "A process $X(t)$ is said to be *stationary in the wide sense* if its mean is a constant and the autocorrelation function depends only on the time difference" [57, p. 136].

Bibliography

1. Atkinson, Kendall E. *An Introduction to Numerical Analysis* (2nd Edition). New York: Wiley, 1989.

2. Berger, Toby and Donald W. Tufts. "Optimal Pulse Amplitude Modulation Part I: Transmitter–Receiver Design and Bounds from Information Theory," *IEEE Transactions on Information Theory*, 196–208 (April 1967).

3. Biglieri, Ezio, et al. "The Optimal Linear Receiving Filter for Digital Transmissions Over Nonlinear Channels," *IEEE Transactions on Information Theory*, *35*(3):620–625 (May 1989).

4. Blackman, R. B. and J. W. Tukey. *The Measurement of Power Spectra from the Point of View of Communications Engineering*. New York: Dover Publications, 1958.

5. Boucher, Ronald E. and Joseph C. Hassab. "Analysis of Discrete Implementation of Generalized Cross Correlator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *ASSP-29*(3):609–611 (June 1981).

6. Carter, G. Clifford, et al. "The Smoothed Coherence Transform," *Proceedings of the IEEE*, 1497–1498 (October 1973).

7. Casella, George and Roger L. Berger. *Statistical Inference* (2nd Edition). Pacific Grove, California: Duxbury, 2002.

8. Cohen, Leon. *Time-Frequency Analysis*. Upper Saddle River, New Jersey: Prentice Hall, 1995.

9. D'Andrea, Aldo N., et al. "The Modified Cramer–Rao Bound and Its Application to Synchronization Problems," *IEEE Transactions on Communications*, *42*:1391–1399 (February/March/April 1994).

10. Dillard, Robin A. "Detectability of Spread-Spectrum Signals," *IEEE Transactions on Aerospace and Electronic Systems*, *AES-15*(4):526–537 (July 1979).

11. Dwork, B. M. "Detection of a pulse Superimposed on Fluctuation Noise," *Proceedings of the IRE*, *38*:771–774 (July 1950).

12. Ericson, Thomas. "Structure of Optimum Receiving Filters in Data Transmission Systems," *IEEE Transactions on Information Theory*, 352–353 (May 1971).

13. Ferguson, Thomas S. *Mathematical Statistics: A Decision Theoretic Approach*. 111 Fifth Avenue, New York, New York 10003: Academic Press Inc., 1967.

14. Fong, G., et al. "An Algorithm for Improved Signal-Selective Time-Difference Estimation for Cyclostationary Signals," *IEEE Signal Processing Letters*, *1*:38–40 (February 1994).

15. Gardner, William A. "Structural Characterization of Locally Optimum Detectors in Terms of Locally Optimum Estimators and Correlators," *IEEE Transactions on Information Theory*, *IT-28*(6):924–932 (November 1982).

16. Gardner, William A. "Spectral Correlation of Modulated Signals: Part I—Analog Modulation," *IEEE Transactions on Communications*, *COM-35*(6):584–594 (June 1987).

17. Gardner, William A. "Signal Interception: a Unifying Theoretical Framework for Feature Detection," *IEEE Transactions on Communications*, *36*:897–906 (August 1988).

18. Gardner, William A. *Statistical Spectral Analysis: A Nonprobabilistic Theory*. Englewood Cliffs, New Jersey: Prentice Hall, 1988.

19. Gardner, William A. "Exploitation of Spectral Redundancy in Cyclostationary Signals," *IEEE SP Magazine*, 14–36 (April 1991).

20. Gardner, William A. "Two Alternative Philosophies for Estimation of the Parameters of Time-Series," *IEEE Transactions on Information Theory*, *37*:216–218 (January 1991).

21. Gardner, William A., et al. "Spectral Correlation of Modulated Signals: Part II—Digital Modulation," *IEEE Transactions on Communications*, *COM-35*(6):595–601 (June 1987).

22. Gardner, William A. and Chih-Kang Chen. "Selective Source Location by Exploitation of Spectral Coherence." *Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling*. 274–279. 1988.

23. Gardner, William A. and Chih-Kang Chen. "Signal-Selective Time-Difference-of-Arrival Estimation for Passive Location of Man-Made Signal Sources in Highly Corruptive Environments, Part I: Theory and Method," *IEEE Transactions on Signal Processing*, *40*(5):1168–1184 (May 1992).

24. Gardner, William A. and Chih-Kang Chen. "Signal-Selective Time-Difference-of-Arrival Estimation for Passive Location of Man-Made Signal Sources in Highly Corruptive Environments, Part II: Algorithms and Performance," *IEEE Transactions on Signal Processing*, *40*(5):1185–1197 (May 1992).

25. George, Donald A., et al. "An Adaptive Decision Feedback Equalizer," *IEEE Transactions on Communication Technology*, *COM-19*(3):281–293 (June 1971).

26. Gisselquist, Daniel E. "Minimum Mean Square Error Filters for Binary Phase Shift Keyed Signals." *2003 IEEE Military Communications Conference, MILCOM '03*. 2003.

27. Gisselquist, Daniel E. "Predicting Binary Phase Shift Keyed Demodulator Performance." *Proceedings of the Thirty–Seventh Annual Asilomar Conference on Signals, Systems, and Computers*. November 2003.

28. Gitlin, R. D., et al. *Data Communication Principles*. Plenum Press, 1992.

29. Hahn, William R. and Steven A. Tretter. "Optimum Processing for Delay-Vector Estimation in Passive Signal Arrays," *IEEE Transactions on Information Theory*, *IT-19*(5) (September 1973).

30. Haykin, Simon. *Adaptive Filter Theory* (3rd Edition). Upper Saddle River, New Jersey: Prentice Hall, 1996.

31. Hoffman, Kenneth and Ray Kunze. *Linear Algebra* (2nd Edition). New Jersey: Prentice Hall, 1971.

32. Hurd, H. L. "Stationarizing properties of random shifts," *SIAM J. Appl. Math.*, *26*(1):203–211 (1974).

33. Jacob, Bill. *Linear Algebra*. New York: W. H. Freeman and Company, 1990.

34. Jenq, Yih-Chyun, et al. "Probability of Error in PAM Systems with Intersymbol Interference and Additive Noise," *IEEE Transactions on Information Theory*, *IT-23* (September 1977).

35. Johnson, Don H. and Dan E. Dudgeon. *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, New Jersey: Prentice Hall, 1993.

36. Kahn, David. *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet*. New York, NY: Scribner, 1996.

37. Knapp, Charles H. and G. Clifford Carter. "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *ASSP-24*(4):320–327 (August 1976).

38. Laha, R. G. and V. K. Rohatgi. *Probability Theory*. New York: John Wiley and Sons, 1979.

39. Laurent, Pierre A. "Exact and Approximate Construction of Digital Phase Modulations by Superposition of Amplitude Modulated Pulses (AMP)," *IEEE Transactions on Communications*, *COM-34*(2):150–160 (February 1986).

40. Lucky, R. W. "Techniques for Adaptive Equalization of Digital Communications Systems," *The Bell System Technical Journal*, *45* (April 1966).

41. Martinez, Peter. "PSK31: A New Radio–Teletype Mode," *RadCom* (December 1998).

42. Mengali, Umberto and Michele Morelli. "Decomposition of M–ary CPM Signals into PAM Waveforms," *IEEE Transactions on Information Theory*, *41*(5):1265–1275 (September 1995).

43. Mills, R. F. and G. E. Prescott. "A Comparison of Various Radiometer Detection Models," *IEEE Transactions on Aerospace and Electronic Systems*, *32*:467–473 (January 1996).

44. Napolitano, Antonio and Chad M. Spooner. "Cyclic Spectral Analysis of Continuous–Phase Modulated Signals," *IEEE Transactions on Signal Processing*, *49*:30–44 (January 2001).

45. Nemsick, L. W. and E. Geraniotis. "Adaptive Multichannel Detection of Frequency-Hopping Signals," *IEEE Transactions on Communications*, *40*:1502–1511 (September 1992).

46. Nyquist, H. "Certain Topics in Telegraph Transmission Theory," *Transactions of the AIEE (Communications and Electronics)*, *47*:617–644 (April 1928).

47. Oppenheim, Alan V. and Ronald W. Schafer. *Discrete–Time Signal Processing*. Englewood Cliffs, New Jersey: Prentice Hall, 1989.

48. Press, William H., et al. *Numerical Recipes in C: The Art of Scientific Computing* (2nd Edition). New York: Cambridge Press, 1992.

49. Priestley, M. B. *Spectral Analysis and Time Series*. San Diego, CA: Academic Press, 1981.

50. Roberts, Richard A. and Clifford T. Mullis. *Digital Signal Processing*. Reading, Massachusetts: Addison-Wesley, 1987.

51. Rostaing, P., et al. "Cyclic Detection in a Nonwhite Gaussian Noise." *1995 International Conference on Acoustics, Speech, and Signal Processing 3*. 2008–2011. 1995.

52. Roth, Peter R. "Effective Measurements using Digital Signal Analysis," *IEEE Spectrum*, 62–70 (April 1971).

53. Scharf, Louis L. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, Massachusetts: Addison–Wesley Publishing Company, 1991.

54. Scharf, Louis L. and Benjamin Friedlander. "Matched Subspace Detectors," *IEEE Transactions on Signal Processing*, *42*(8):2146–2157 (August 1994).

55. Schell, S. V. Personal Correspondence.

56. Schell, S. V. "Corollaries, for Cyclostationary Signals, to Whittle's Theorem on the Cramér–Rao Bound." Sponsored under NSF grant MIP-95-12732 and U.S. Army Research Office contract DAAH 04-95-1-0328. Acquired through correspondence., January 1998.

57. Shanmugan, K. Sam and A. M. Breipohl. *Random Signals: Detection, Estimation, and Data Analysis*. New York: John Wiley & Sons, 1988.

58. Sklar, Bernard. *Digital Communications: Fundamentals and Applications* (2nd Edition). Upper Saddle River, New Jersey: Prentice Hall, 2001.

59. So, H. C., et al. "Detection of Random Signals via Spectrum Matching," *IEEE Transactions on Aerospace and Electronic Systems*, *38*(1):301–306 (January 2002).

60. Sonnenschein, A. and P. M. Fishman. "Limitations on the Detectability of Spread-Spectrum Signals." *1989 IEEE Military Communications Conference, MILCOM '89: Bridging the Gap. Interoperability, Survivability, Security.2*. 364–369. 1989.

61. Streight, D. A. *Maximum Likelihood Estimates for the Time and Frequency Differences of Arrival of Cyclostationary Digital Communications Signals*. PhD dissertation, Naval Postgraduate School, Monterey, CA, 1999.

62. Streight, David A., et al. "Maximum Likelihood Estimates of the Time and Frequency Differences of Arrival of Weak Cyclostationary Digital Communications Signals." *MILCOM 2000: 21st Century Military Communications Conference Proceedings2*. 957–961. 2000.

63. Stremler, Ferrel G. *Introduction to Communication Systems* (3rd Edition). Reading, Massachusetts: Addison–Wesley Publishing, 1992.

64. Therrien, Charles W. *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs, New Jersey 07632: Prentice-Hall, Inc., 1992.

65. Tufts, Donald W. "Nyquist's Problem—The Joint Optimization of Transmitter and Receiver in Pulse Amplitude Modulation," *Proceedings of the IEEE*, 248–259 (March 1965).

66. Turin, George L. "An Introduction to Matched Filters," *IRE Transactions on Information Theory*, 311–329 (June 1960).

67. Urkowitz, Harry. "Energy Detection of Unknown Deterministic Signals," *Proceedings of the IEEE*, *55*(4):523–531 (April 1967).

68. Van Trees, Harry L. *Detection, Estimation, and Modulation Theory*, *1*. New York: John Wiley & Sons, 1968.

69. Van Trees, Harry L. *Detection, Estimation, and Modulation Theory: Radar–Sonar Signal Processing and Gaussian Signals in Noise*, *3*. New York: John Wiley & Sons, 1968.

70. Widrow, Bernard, et al. "Adaptive Noise Cancelling: Principles and Applications," *Proceedings of the IEEE*, 1692–1716 (December 1975).

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
12-03-2004	**Doctoral Dissertation**	Sep 2001 – Mar 2004

4. TITLE AND SUBTITLE		5a. CONTRACT NUMBER
A LINEAR SUBSPACE APPROACH TO BURST COMMUNICATION SIGNAL PROCESSING		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Gisselquist, Daniel, E., Major, USAF	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	8. PERFORMING ORGANIZATION REPORT NUMBER
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-7765	AFIT/DS/ENG/04-02

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
James P. Stephens and Chahira M. Hopper AFRL/SNRW 2241 Avionics Circle WPAFB, OH 45433-7321 DSN: 785-5579	Thomas E. Brehm NASIC/DES 4180 Watson Way WPAFB, OH 45433-5648 DSN: 784-4668	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This dissertation focuses on the topic of burst signal communications in a high interference environment. It derives new signal processing algorithms from a mathematical linear subspace approach instead of the common stationary or cyclostationary approaches. The research developed new algorithms that have well-known optimality criteria associated with them. The investigation demonstrated a unique class of multisensor filters having a lower mean square error than all other known filters, a maximum likelihood time difference of arrival estimator that outperformed previously optimal estimators, and a signal presence detector having a selectivity unparalleled in burst interference environments. It was further shown that these improvements resulted in a greater ability to communicate, to locate electronic transmitters, and to mitigate the effects of a growing interference environment.

15. SUBJECT TERMS
Burst Communication, Signal Processing, Cyclostationary Signal Processing, Time Difference of Arrival (TDOA), Presence Detection, Cramer-Rao Bound, Demodulation, Filtering

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
REPORT U	**ABSTRACT** U	**c. THIS PAGE** U	UU	273	Donald R. Kitchen, USAF (ENG)
					19b. TELEPHONE NUMBER *(Include area code)* (937) 255-2024; e-mail: Donald.Kitchen@afit.edu